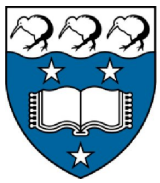
The nervous 90s: A Bayesian analysis of batting in Test cricket

Oliver George Stevenson

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Science in Statistics



Department of Statistics,

The University of Auckland,

2017

Abstract

Cricketing knowledge tells us batting is more difficult early in a player's innings, but gets easier as a player becomes familiar with the local conditions. Using Bayesian inference and nested sampling techniques, a model is developed to predict the Test match batting abilities of international cricketers. The model allows for the quantification of players' initial and equilibrium batting abilities, and the rate of transition between the two. Implementing the model using a hierarchical structure provides more general inference concerning a selected group of international opening batsmen from New Zealand. More complex models are then developed, which are used to identify the presence of any score-based variation in batting ability among a group of modern-day, world-class batsmen. Additionally, the models are used to explore the plausibility of popular cricketing superstitions, such as the 'nervous 90s'. Evidence is found to support the existence of score-based variation in batting ability, however there is little support to confirm a widespread presence of the 'nervous 90s' affecting player batting ability. Practical implications of the findings are discussed in the context of specific match scenarios.

Acknowledgements

Firstly, a huge thanks to my supervisor, Brendon Brewer, for his guidance, patience and impressive cricketing knowledge in the last year; without you this project would not have been possible. It has been a great experience combining two of my favourite things — cricket and statistics — and coming out the other end with both a publication and Masters degree. I have learnt a lot from you and look forward to continuing to work alongside you in the near future.

I also wish to thank my parents, Helen and Craig, for their support during the year. Thank you to my brother Ben, I truly appreciate the amount of effort you put in to answering all of my questions, no matter how busy you were. Finally, thanks to my sister Kate, my girlfriend Steph, Chris and Charlotte for all the fun times had throughout the past year.

Contents

Chapter 1

Introduction

Since the inception of statistical record-keeping in cricket, a player's batting ability has primarily been recognised using a single number, their batting average. However, in cricketing circles it is common knowledge that a player will not begin an innings batting to the best of their ability. Rather, it takes time to adjust both physically and mentally to the specific match conditions. This process is nicknamed 'getting your eye-in'. External factors such as the weather and the state of the pitch are rarely the same in any two matches and can take time to get used to. Additionally, batsmen (a term commonly used to refer to both male and female cricketers) will often arrive at the crease with the match poised in a different situation to their previous innings, requiring a different mental approach. Subsequently, batsmen are regularly seen to be dismissed early in their innings while still familiarising themselves with the specific match conditions. This suggests that a constant-hazard model, whereby the probability of a batsman being dismissed on their current score (called the *hazard*) remains constant regardless of their score, is not ideal for predicting when a batsman will get out.

Compared with many sports, cricket is unique in the sense that physical differences between matches, such as the weather and the pitch, have a significant bearing on how a particular match will be played out. Vastly different approaches to the game are seen between the dusty, spinner-friendly pitches of India, and the green seaming pitches

common in the likes of England and New Zealand. As these external factors vary considerably between matches, batsmen must adapt their technique and game plan accordingly to best cope with the local conditions, which can be difficult, especially in foreign environments.

Given the statistical culture that has developed with the growth of cricket, statistics such as batting averages have become the acknowledged method of best judging a player's ability. Other statistics, such as the number of runs a player has scored in their career, or the number of occasions they have passed significant milestones such as 50 and 100 are also useful. However, coaches, commentators and players alike, can all get a quick, and often fairly accurate understanding of an individual's batting ability, simply by looking at their batting average. For a sport as complex as cricket, where a single match can continue for up to five days and a career can span over 20 years, it seems inadequate to measure a player's batting ability over the course of an innings, using just one number.

It would be of practical use to both coaches and players to have a more flexible method of quantifying how well a batsman is performing at any given stage of their innings. Identifying players' batting weaknesses and improving team selection can be aided by tools that estimate measures such as (1) how well batsmen perform when they first arrive at the crease, (2) how much better they perform once they have got their 'eye-in' and (3) how long it takes them to accomplish this.

Furthermore, due to the statistical nature of the game, milestones can also play a large role in a batsman's innings. Passing scores of significance, such as 50 and 100, carries a mark of distinction among cricketers and doing so becomes a permanent part of a player's career record. Psychological studies have indicated mood can have a significant impact on a cricketer's performance (Totterdell, 1999), suggesting player concentration levels may change over the course of an innings. As a result, it is not uncommon to see batsmen lose concentration after passing significant scores and playing risky shots they may not have otherwise attempted. Superstition also has a place among the hearts of many cricketers, which may result in a lapse (or perhaps an increase) in judgement and concentration when nearing so-called 'unlucky numbers'.

The term 'nervous 90s' has been coined to refer to a form of analysis paralysis suffered by a batsman who is currently on a score between 90 and 99 (i.e. near the significant milestone score of 100). The 'nervous 90s' are a favourite among the likes of commentators and the media, who will attribute almost any dismissal in the 90s, due to nerves. A player may bat more conservatively than they might otherwise while in the 90s, which may be indicative of the fact they are aware how close they are to scoring a century. Opposition captains may use this as an opportunity to set more attacking fields to a batsman in the 'nervous 90s', in the hopes of creating additional pressure and inducing the batsman into a false stroke. However, despite plenty of anecdotal evidence that many players do become more nervous while on scores in the 90s, there is no clear evidence that these nerves adversely affect batting ability.

Player mood and concentration may also be affected by various off-field anomalies. South African batsman Neil McKenzie was the culprit of one of the more bizarre superstitions, always taping his bat to the dressing-room ceiling before going out to bat, while all-time Indian great, Sachin Tendulkar, was known for always strapping on his left pad before his right. It may be that such players, with strange pre-match rituals, are more likely to succumb to superstitions such as the 'nervous 90s'. Therefore, accounting for these additional factors, one might expect to see deviations in a batsman's ability around certain scores, rather than reaching a plateau at some peak ability.

1.1 Studies in cricket

Statistical analysis is particularly valid for sports such as cricket, given the closed nature of the skills of batting and bowling. Unlike sports such as rugby and football, this generally allows the large amounts of data collected each match, to be treated independently of the specific match scenario. Therefore, due to the data-rich nature of the sport, cricket has been the focus of numerous statistical studies.

From a public perspective, it is difficult to know exactly what international and domestic teams focus on in terms of cricketing statistics, as the information is highly

sensitive. However, despite the abundance of data available in cricket, it is yet to be commercially exploited to the same extent as its American cousin, baseball. Much of the 'statistical' analysis performed in the public realm, outside academia, is relatively low level, often revolving around simple summary statistics and ground histories.

In the past, having a statistical analyst as part of a team's coaching staff would be seen as a waste of time and money, however with the resources and computing power available today, we can only assume having a dedicated statistician is becoming become more and more of a necessity. Consequently, the volume of cricket-related research has grown since the turn of the century, with studies tending to fall in one of four categories

1. Achieving a fair result in interrupted matches.
2. Predicting the outcome of a match.
3. Optimising playing strategies.
4. Analysing player performance and ability.

Achieving a fair result

Many of the early studies focussing on cricket fall into the first category, as statisticians trialled various methods of resolving interrupted (usually weather-related) limited-overs cricket matches. In the late 1990s the Duckworth-Lewis method (Duckworth & Lewis, 1998) was developed and has since become the entrenched method of dealing with interrupted matches.

Since its implementation, the Duckworth-Lewis method (now Duckworth-Lewis-Stern or D/L/S) has become the most well-known statistical tool used in cricket. Various attempts have been made to fine-tune or better the Duckworth-Lewis method (Thomas, 2002; Jayadevan, 2002; Carter & Guthrie, 2004), however none have succeeded in supplanting it as the international cricketing standard.

Predicting the outcome of a match

Following the development of the Duckworth-Lewis method, cricketing research has recently shifted to have more of a prediction and analytical focus. As tends to happen when sport and statistics collide, outcome prediction has been scrutinised heavily by coaches, bookies, punters and spectators alike. The WASP tool (Winning And Score Predictor) (Brooker & Hogan, 2011) has enjoyed some public success, achieving somewhat of a cult status in parts of the cricketing world, featuring on televised coverage of Twenty20 and one-day matches in New Zealand and England. WASP uses historical and local ground data to predict the score to be made by the team batting first, followed by the probability of the team batting second chasing the score down. However, despite the attempts of cricket-mad statisticians (Brooks et al., 2002; Bailey & Clarke, 2006; Swartz et al., 2009; Brooker & Hogan, 2011), as with all sports, no sure-fire method has emerged for determining which side will come away victorious.

Optimising playing strategies

Measures such as batting and bowling averages have also been used to optimise both player and team performance, in order to fine-tune both playing strategies (Clarke, 1988; Clarke & Norman, 1999; Preston & Thomas, 2000; Davis et al., 2015), and decision making (Clarke & Norman, 2003; Swartz et al., 2006; Norman & Clarke, 2010) during a match.

Several of the developed methods in this area, while interesting, are of limited use, as they tend to focus on very specific match circumstances, making it difficult to apply them in a broader scope. Such strategies would be useful for teams if they could be applied concurrently, as a match is played.

Analysing player performance and ability

The content in this thesis falls into the category of analysing player ability and focusses on developing new player performance measures, specifically to assess how batting

abilities of players change during an innings. Surprisingly few studies have explored possible player performance measures which better explain batting ability than the humble batting average.

One of the first documented cases of statistics being used to model batting scores occurred in the pre-computing era; Elderton & Wood (1945) provided empirical evidence to support the claim that a batsman's scores could be modelled using a geometric progression. However, the geometric assumption does not necessarily hold for all players (Kimber & Hansford, 1993), namely due to its difficulty in fitting the inflated number of scores of 0 appearing in many players' career records. To account for this, Bracewell & Ruggiero (2009) proposed to model player batting scores using a distribution called the 'Ducks 'n' runs' distribution, which is a mixture of a beta distribution and a geometric distribution. The beta distribution is used to model player contribution, defined as the proportion of runs an invididual contributes towards their team's total. As the probability of having a contribution of zero is equal to 0 under the beta distribution, a small continuity correction is applied. This allows for the calculation of a player's probabilty of failing to contribute any runs (ducks) towards their team's total. The geometric component then describes the distribution of non-zero scores (runs).

Rather than model batting scores, Kimber & Hansford (1993) used nonparametric models to derive a player's hazard at a given score, estimating dismissal probabilities as a batsman's innings progresses. Methods for estimating the hazard function for discrete and ordinal data have long-existed in survival analysis (McCullagh, 1980; Allison, 1982), and have applications across a wide range of disciplines. However, the present case may be considered unusual in the context of discrete hazard functions, given the large number of ordered, discrete points (i.e. number of runs scored) (Agresti & Kateri, 2011). Estimating the hazard function allows us to observe how a player's dismissal probability (and therefore, batting ability) varies over the course of their innings. While Kimber & Hansford (1993) found batsmen were more likely to get out early in their innings, due to the sparsity of data at higher scores these estimates quickly become unreliable and the estimated hazard function jumps er-

ratically between scores. Cai et al. (2002) addressed this issue using a parametric smoother on the hazard function, however given the underlying function they used is still a nonparametric estimator, the problem of data sparsity still remains an issue and continues to distort the hazard function at higher scores.

Bayesian stochastic methods have also been used to measure batting performance (Koulis et al., 2014; Damodaran, 2006). Koulis et al. (2014) proposed a model for evaluating performance based on player form, however this only allows for innings to innings comparisons in terms of batting ability, rather than comparisons *during* an innings. On the other hand, Damodaran (2006) provides a method which does allow for within-innings comparisons, but lacks a natural cricketing interpretation. Various other performance metrics have been proposed, however these have been developed in relation to limited overs cricket (Lemmer, 2004, 2011; Damodaran, 2006; Koulis et al., 2014). Our focus is exclusively on Test and first-class cricket, as limited overs cricket introduces a number of complications (Davis et al., 2015).

As an alternative, Brewer (2008) proposed a Bayesian parametric model to estimate a player's current batting ability (via the hazard function) given their current score, using a single change-point model. This allows for a smooth transition in the hazard between a batsman's 'initial' and 'eye-in' states, rather than the sudden jumps seen in Kimber & Hansford (1993) and to an extent Cai et al. (2002). Based on our knowledge of cricket, it is fair to assume that batsmen are more susceptible early in their innings and tend to perform better as they score more runs. The findings from Brewer (2008) confirm these assumptions, however, of particular note, was that the batsmen with the highest career batting averages, are not necessarily the best batsmen when beginning an innings.

A primary aim of this thesis was to further develop the model in Brewer (2008) to better identify how a batsman's ability changes over the course of an innings, ideally giving a better indication of batting ability than their batting average. The resulting models have practical implications in terms of how they can be applied in a match situation, and have the added benefit of a natural cricketing interpretation, which coaches and players can easily understand.

1.2 Bayesian inference

As the foundation of our models is based on the approach of Brewer (2008), those detailed in this thesis were developed within a Bayesian framework. Working under the Bayesian paradigm, as opposed to the more traditional frequentist paradigm, allows us to to express all forms of uncertainty in terms of probability, providing us with the tools to update our beliefs in the face of new data (O'Hagan & Forster, 2004).

Applying the principles of Bayesian inference to a problem, firstly requires choices to be made regarding the questions we want to answer and the assumptions we are willing to make (Brewer, 2014). Under a Bayesian approach, part of this initial decision making process requires us to specify *prior* distributions for our parameters of interest, θ. These prior distributions represent our initial state of knowledge regarding the parameters and can be written as $p(\theta|m)$, where m are our model assumptions. Working within a Bayesian context with cricketing data is convenient, as we already have considerable knowledge on how we expect the game to be played. As such, we can assign appropriate subjective prior distributions to our parameters of interest, reflecting our knowledge of cricket.

Our goal is to make meaningful inference regarding our parameters, θ, by inferring from the data, d. Upon observing the data, we are able to update our prior beliefs regarding θ, expressing them as a posterior distribution $p(\theta|d, m)$. However, in order to get from the prior distribution to the posterior distribution, we must also consider what the likelihood of observing our data was, given our prior beliefs, i.e. $p(d|\theta, m)$. If the data we observe is drastically different from what we expected under our prior assumptions for θ, we must update our posterior beliefs accordingly, so that future data of the same kind is more plausible.

The equation connecting the prior and posterior distribution for θ, is known as Bayes' theorem. Considering two propositions, A and B, we can derive Bayes' theorem

using the product rule

$$\begin{aligned} P(A \cap B) &= P(A)\ P(B|A) \\ P(B \cap A) &= P(B)\ P(A|B) \\ \therefore P(A|B) &= \frac{P(A)\ P(B|A)}{P(B)}. \end{aligned} \tag{1.1}$$

Therefore, we can use Bayes' theorem from Equation 1.1 to express the posterior distribution for θ as

$$p(\theta|d, m) = \frac{p(\theta|m)\ p(d|\theta, m)}{p(d|m)}, \tag{1.2}$$

where $p(d|m)$ is the marginal likelihood or 'evidence'. This is the probability of observing the data, computed by integrating across all possible values of θ, weighted by our prior beliefs for each particular value of θ. As it can be difficult to obtain $p(d|m)$, we can simplify the posterior distribution to

$$p(\theta|d, m) \propto p(\theta|m)\ p(d|\theta, m). \tag{1.3}$$

Since we are often dealing with multiple parameters, the posterior distribution is often of high dimension, which can be difficult to express numerically. Therefore, in order to quantify our beliefs regarding θ once we have observed the data, we must summarise the posterior distribution. This can be achieved using numerical techniques such as Markov Chain Monte Carlo (MCMC) methods, which samples from the posterior distribution. Popular MCMC methods include the Metropolis-Hastings algorithm (Hastings, 1970), and Gibbs sampling (Geman & Geman, 1984).

1.2.1 Nested sampling

Nested sampling is a technique developed by physicist John Skilling (Skilling, 2006), that enables the computation of a posterior distribution. More specifically, nested sampling estimates directly how the likelihood function relates to the prior mass. A

major advantage of the nested sampling algorithm is that the primary result is the marginal likelihood, or evidence, a quantity which cannot be easily computed using standard MCMC methods.

Formally, nested sampling computes

$$Z = \text{marginal likelihood} = \int p(\theta)\ L(\theta)\ d\theta,$$

where $p(\theta)$ is the prior distribution for θ and $L(\theta)$ is the likelihood function. Therefore, if d are our data and m are our background model assumptions, the evidence quantities, Z, are the probabilities of observing the data given our model assumptions, i.e. $P(d|m)$. The marginal posterior distributions for θ, are also readily available from the nested sampling computation, by taking weighted samples of θ from the nested sampling run (Skilling, 2006).

Consider two models, with unique assumptions m_1 and m_2 and parameters θ_1 and θ_2. Using Bayes theorem (Equation 1.1) we can calculate the posterior distributions for θ_1 and θ_2 when fitting each model to identical data. The models can then be compared, using the evidence ratio or Bayes factor, which is the Bayesian equivalent to the likelihood ratio (Kass & Raftery, 1995)

$$\frac{P(m_1|d)}{P(m_2|d)} = \frac{P(m_1)}{P(m_2)} \times \frac{P(d|m_1)}{P(d|m_2)}. \tag{1.4}$$

Equation 1.4 calculates the posterior odds, which is simply the prior odds multiplied by Bayes factors. Where standard MCMC methods only compute the posterior $p(\theta|d, m)$, model comparison becomes trivial if competing models are fitted using nested sampling, as the marginal likelihood is a primary result of the computation.

Diffusive nested sampling

The simple implementation of the nested sampling algorithm given by Skilling (2006) is sufficient when dealing with models of relatively low dimensionality. However, in cases where the number of model parameters is large, standard nested sampling can

become computationally impractical. Additionally, nested sampling can struggle in cases where the likelihood function exhibits multimodality, usually getting stuck at one of the local maximum likelihood values.

Diffusive nested sampling (DNS) (Brewer et al., 2011; Brewer & Foreman-Mackey, 2016) is a variant of the nested sampling algorithm, which can be applied in scenarios where standard nested sampling falls short. In high dimensional problems DNS produces more accurate estimates for the marginal likelihood, Z, and ought to outperform classic nested sampling on multimodal problems (Brewer & Foreman-Mackey, 2016).

Each of the models detailed in this thesis were fitted using either classic nested sampling (Skilling, 2006) or diffusive nested sampling (Brewer et al., 2011; Brewer & Foreman-Mackey, 2016). Those which used classic nested sampling were implemented using the Julia programming language[1] (Bezanson et al., 2014), while models which required diffusive nested sampling, used a C++ (ISO, 2012) implementation of the algorithm that calls Julia to evaluate the likelihood function. The post-processing and manipulation of data was primarily dealt with using R (R Core Team, 2015), as well as the construction of most graphics (Wickham, 2009).

1.3 The present study

This thesis focusses on the fitting and analysis of models which describe a cricket player's batting ability over the course of an innings. A class of flexible models are investigated to determine whether or not we can tentatively confirm or deny the effects of popular cricketing superstitions, such as the 'nervous 90s'.

In Chapter 2, we propose an alternative Bayesian model to Brewer (2008) for inferring a batsman's hazard function from their career batting record. This model is then applied as part of a hierarchical inference in Chapter 3, allowing us to make generalised statements about a wider group of players (in this case, opening batsmen who have represented New Zealand), rather than being restricted to analysing a single

[1] https://github.com/eggplantbren/NestedSampling.jl

player at a time. Both Chapters 2 and 3 provide the basis of publication for Stevenson & Brewer (2017).

Using the initial model detailed in Chapter 2 as a foundation, more flexible models are explored and developed in Chapter 4. These models allow for increased temporal variation in player batting ability across a player's innings, and are used to evaluate the existence of any detrimental effects on batting ability due to the 'nervous 90s'.

Chapter 5 uses the marginal likelihood values computed for each model to perform model comparison. The thesis is then summarised using a single marginal likelihood value, which allows for the present set of models to be compared with future models, if the same data set is applied.

Finally, Chapter 6 summarises the findings for each of the proposed models, suggesting practical uses that could be implemented by teams and coaches.

Chapter 2

The exponential varying-hazard model

2.1 Overview

This chapter details the initial varying-hazard model, presented in Stevenson & Brewer (2017), used to model batting ability over the course of a player's innings (hereafter referred to as the *exponential varying-hazard model*). The model likelihood is defined in Section 2.2, with particular emphasis placed on the parameterisation of the hazard function, outlined in Section 2.2.1.

Once the model is sufficiently defined, its performance is compared with the model of Brewer (2008), analysing the same data set of retired international cricketers (Section 2.3.2). The model is implemented using the Julia programming language (Bezanson et al., 2014) and allows us to quantify each players'

1. Ability when they first arrive at the crease.
2. Ability when they have their 'eye-in'.
3. The speed of the transition between these two states.

The conclusions drawn from the model fitting process are similar to those in Brewer (2008), though differing levels of uncertainty in parameter estimates are ob-

tained. A clear difference in initial and 'eye-in' ability is observed for most players, validating our belief that a constant hazard model is not usually ideal for predicting when a batsman will be dismissed.

2.2 Model structure

The derivation of the model likelihood follows the method detailed in Brewer (2008) and Stevenson & Brewer (2017). In cricket, a player bats and continues to score runs until (1) they are dismissed, (2) every other player in his team is dismissed, (3) his team's innings is concluded via a declaration or (4) the match ends. Consider the score $X \in \{0, 1, 2, 3, ...\}$ that a batsman scores in a particular innings. Define the hazard function, $H(x) \in [0, 1]$, as the probability the batsman gets out on score x, given they are currently on score x; i.e., the probability the batsman scores no more runs

$$H(x) = P(X = x|X \geq x) = \frac{P(X = x, X \geq x)}{P(X \geq x)} = \frac{P(X = x)}{P(X \geq x)}. \tag{2.1}$$

Throughout this section, all probabilities and distributions are conditional on some set of parameters, θ, which will determine the form of $H(x)$ and therefore $P(X = x)$. We proceed by defining $G(x) = P(X \geq x)$ as the 'backwards' cumulative distribution. Using this definition, Equation 2.1 can be written as a difference equation for $G(x)$

$$\begin{aligned} G(x) &= P(X \geq x) \\ G(x) &= P(X = x) + P(X \geq x + 1) \\ G(x) &= H(x)G(x) + G(x + 1) \\ G(x + 1) &= G(x) - H(x)G(x) \\ G(x + 1) &= G(x)[1 - H(x)]. \end{aligned} \tag{2.2}$$

With the initial condition $G(0) = 1$ and an assumed functional form for $H(x)$, we

can calculate $G(x)$ for $x > 0$:

$$G(x) = \prod_{a=0}^{x-1}[1 - H(a)]. \tag{2.3}$$

This is the probability of scoring one run, times the probability of scoring two runs given the batsman scored one run, etc., up to the probability of scoring x runs given that the batsman scored $x - 1$ runs. Therefore, the probability mass function for X is given by the probability of surviving up until score x, then being dismissed:

$$P(X = x) = H(x) \prod_{a=0}^{x-1} [1 - H(a)] \,, \tag{2.4}$$

which is the probability distribution for the score in a single innings, given a model of H.

When we infer the parameters θ from data, this expression provides the likelihood function. For multiple innings we assume conditional independence, and for not out innings we use $P(X \geq x)$ as the likelihood, rather than $P(X = x)$. This assumes that for not out scores, the batsman would have gone on to score some unobserved score, conditional on their current score and their assumed hazard function. If we considered these unobserved scores as additional unknown parameters and marginalised them out, we would achieve the same results but at higher computational cost. Thus, if I is the total number of innings and N is the number of not out scores, the probability distribution for a set of conditionally independent scores $\{x_i\}_{i=1}^{I-N}$ and not out scores $\{y_i\}_{i=1}^{N}$ is

$$p(\{x\}, \{y\}) = \prod_{i=1}^{I-N} \Big(H(x_i) \prod_{a=0}^{x_i-1} [1 - H(a)] \Big) \times \prod_{i=1}^{N} \Big(\prod_{a=0}^{y_i-1} [1 - H(a)] \Big). \tag{2.5}$$

When data $\{x, y\}$ are fixed and known, Equation 2.5 above gives the likelihood

for any proposed model of $H(x;\theta)$, the hazard function. The log-likelihood is

$$\log\left[L(\theta)\right] = \sum_{i=1}^{I-N} \log\, H(x_i) + \sum_{i=1}^{I-N}\sum_{a=0}^{x_i-1} \log[1-H(a)] + \sum_{i=1}^{N}\sum_{a=0}^{y_i-1} \log[1-H(a)] \qquad (2.6)$$

where θ is the set of parameters controlling the form of $H(x)$.

2.2.1 Parameterising the hazard function

The parameterisation of the hazard function, $H(x)$, will influence how well we can fit the data, as well as what we can learn from doing so. In order to accurately reflect our belief that batsmen are more susceptible to being dismissed early in their innings, the hazard function should be higher for low values of x (i.e. low scores) and decrease as x increases, as the batsman scores more runs and gets used to the specific match conditions.

Consider the constant hazard model $H(x) = h$, for all scores x, whereby a batsman has equal probability of being dismissed on every score. Deriving the sampling distribution $P(X = x)$ for the constant hazard model (see Equation 2.7) gives the geometric distribution, similar to the approach used by Elderton & Wood (1945).

$$\begin{aligned} H(x) &= P(X = x | X \geq x), \text{from Equation 2.1} \\ P(X = x) &= H(x)\prod_{a=0}^{x-1}\left[1-H(a)\right], \text{from Equation 2.4} \\ &= h\prod_{a=0}^{x-1}\left[1-h\right] \\ &= h(1-h)^x. \end{aligned} \qquad (2.7)$$

Thinking in terms of the geometric distribution, a batsman's expected score is $\mu = \frac{1}{h} - 1$. If we continue to think in terms of the expected number of runs scored by a batsman, μ, it makes sense to parameterise the hazard function in terms of an 'effective batting average', $\mu(x)$, which evolves with score as a batsman 'gets their

eye-in'. This allows us to think of batting ability in terms of batting averages rather than dismissal probabilities, which has a more natural interpretation to the everyday cricketer and non-cricketer alike. We can obtain $H(x)$ from $\mu(x)$ as

$$H(x) = \frac{1}{\mu(x) + 1}. \tag{2.8}$$

Therefore the hazard function, $H(x)$, relies on our parameterisation of a player's effective batting average, $\mu(x)$. It is reasonable to consider that batsmen begin their innings playing with some initial batting ability $\mu(0) = \mu_1$, which increases with the number of runs scored until a peak batting ability μ_2 is reached. Brewer (2008) used a sigmoidal model for the transition from μ_1 to μ_2. However, it is both simpler and probably more realistic to adopt a functional form for $\mu(x)$, where the transition from μ_1 to μ_2 necessarily begins immediately, and where $\mu(0) = \mu_1$ by definition. Therefore we adopt an exponential model, where $\mu(x)$ begins at μ_1 and approaches μ_2 as follows:

$$\mu(x; \mu_1, \mu_2, L) = \mu_2 + (\mu_1 - \mu_2) \exp\left(-\frac{x}{L}\right). \tag{2.9}$$

Our model contains just three parameters: μ_1 and μ_2, the initial and equilibrium batting abilities of the player, and L, the timescale of the transition between these states. Formally, L is the e-folding time and can be understood by analogy with a 'half-life', signifying the number of runs to be scored for 63% of the transition between μ_1 and μ_2 to take place. The major change between the present model and that of Brewer (2008) is that we use just a single parameter, L, to describe the transition between the two effective average parameters, and that μ_1 has a natural interpretation since it equals $\mu(0)$ (i.e. the batsman's initial ability before scoring any runs).

Since we do not expect a batsman's ability to decrease once arriving at the crease, we impose the constraint $\mu_1 \leq \mu_2$. However, it is worth noting there are various instances during a Test match where this assumption may be violated. Batting often becomes more difficult due to a deterioration in physical conditions such as the pitch or light. The introduction of a new bowler or new type of bowler (e.g. a spin rather than seam bowler) may also disrupt the flow of a batsman's innings, especially when

the change coincides with the bowling side opting to take the new ball after 80 (or more) overs. Moreover, batsmen are likely to take some time re-adjusting to the conditions after a lengthy break in play, particularly when resuming their innings at the start of a new day. However, data on these possible confounders is difficult to obtain and it is not clear that including them in the model and then integrating over their related parameters would lead to a large difference from our current approach of ignoring these effects because we do not have the relevant data.

Additionally, we do not expect the transition between the two batting states to be any larger than the player's 'eye-in' effective batting average, so we also impose the restriction $L \leq \mu_2$. To implement these constraints, we performed the inference by re-parameterising from (μ_1, μ_2, L) to (C, μ_2, D) such that $\mu_1 = C\mu_2$ and $L = D\mu_2$, where C and D are restricted to the interval [0, 1]. In terms of the three parameters (C, μ_2, D), the effective average function is

$$\mu(x; C, \mu_2, D) = \mu_2 + \mu_2(C-1)\exp\left(-\frac{x}{D\mu_2}\right). \tag{2.10}$$

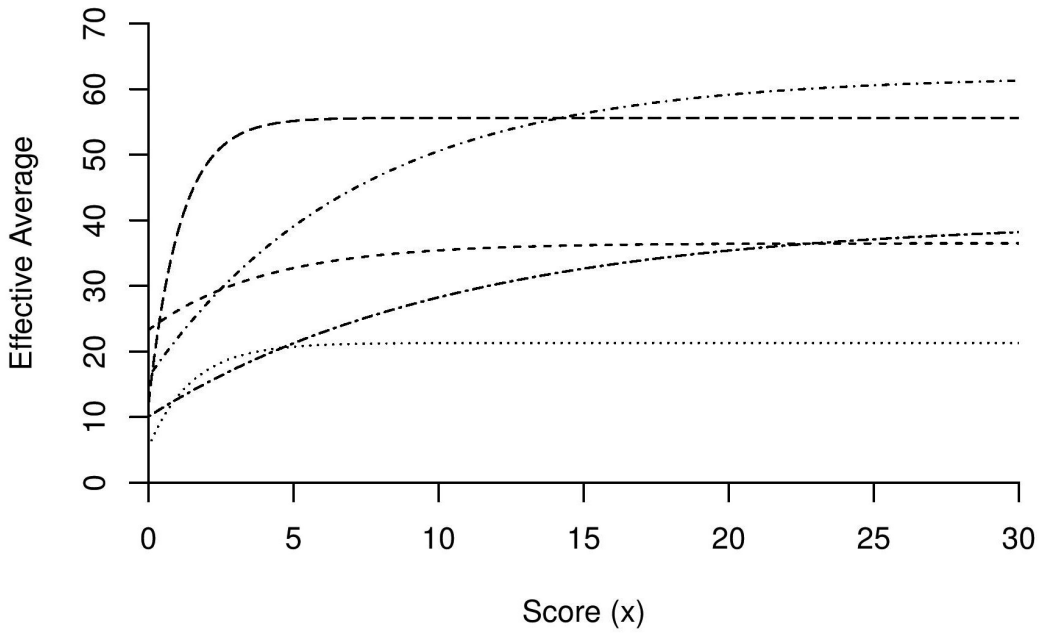


Figure 2.1. Examples of various plausible effective average functions $\mu(x)$, ranging from small to large differences between the initial and equilibrium effective averages μ_1 and μ_2, with both fast and slow transition timescales L.

Therefore the hazard function takes the form

$$H(x) = \frac{1}{\mu_2 + \mu_2(C-1)\exp\left(-\frac{x}{D\mu_2}\right) + 1}. \tag{2.11}$$

See Figure 2.1 for examples of possible effective average functions $\mu(x)$ allowed by this model.

2.2.2 Prior specification

The first stage of the analysis involved evaluating individual player data, using fixed priors for the parameters C, μ_2 and D of each player. This allows us to calculate the joint posterior distributions for μ_1, μ_2 and L for each player. All that is required to analyse individual players using the exponential varying-hazard model is to specify priors on parameters C, μ_2 and D. All parameters are non-negative and C and D lie between 0 and 1.

For μ_2, we selected a prior that loosely coincides with our cricketing knowledge and other anecdotal evidence. A career batting average of 20 is regarded as fairly standard across all cricketers to have played Test match cricket, when considering both batsmen and bowlers. Therefore a Lognormal($\log(25)$, 0.75^2) prior was chosen for μ_2, signifying a prior median 'eye-in' batting average of 25, with a width (standard deviation of $\log(\mu_2)$) of 0.75. The lognormal distribution was preferred as it is a natural and well-known distribution for modelling uncertainty about a positive quantity whose uncertainty spans an order of magnitude or so. This prior implies an expected number of runs per wicket of approximately 33 when batsmen have their 'eye-in', which seems reasonable in the context of Test cricket. The width of 0.75 implies a conservatively wide uncertainty. The prior 68% and 95% credible intervals for μ_2 are $[11.81, 52.93]$ and $[5.75, 108.7]$ respectively.

Selecting a prior which considers a wider range of μ_2 values is ill-advised, as it would allow the model to fit very high 'eye-in' batting abilities for a player with a small sample of high scoring innings. In reality, it is highly improbable that any Test

player will have an effective average greater than 100 at any stage of their innings, except for perhaps the great Sir Donald Bradman, whose cricketing feats are unlikely to be seen again[1].

The priors for C and D were also chosen to be independent from all other parameters in the model. As C and D are restricted to the interval [0, 1], both priors were chosen to follow a beta distribution. The notion of 'getting your eye-in' implies a player's initial batting ability is somewhat worse than their 'eye-in' batting ability. Therefore a Beta(1, 2) prior was assigned to C, emphasising the lower end of the [0, 1] interval, representing a mean initial batting ability that is one-third of a player's 'eye-in' batting ability.

Additionally, we expect a player's e-folding time to be small in comparison to their 'eye-in' batting ability. A Beta (1, 5) prior, further emphasising the lower end of the [0, 1] interval, was assigned to D, representing a mean e-folding time that is one-sixth of a player's 'eye-in' batting ability. If both C and D shared a common Beta(1, 2) prior, the model would favour effective average functions with exceedingly long transition periods between a batsman's initial and 'eye-in' batting abilities.

These priors are presented in Figures 2.2, 2.3 and 2.4, and allow for a range of plausible hazard functions (see Figure 2.1). The overall Bayesian model specification for analysing an individual player using the exponential varying-hazard model is therefore

$$\mu_2 \sim \text{Lognormal}(\log(25), 0.75^2) \tag{2.12}$$

$$C \sim \text{Beta}(1, 2) \tag{2.13}$$

$$D \sim \text{Beta}(1, 5) \tag{2.14}$$

$$\text{log-likelihood} \sim \text{Equation (2.6)} \tag{2.15}$$

[1]Donald Bradman averaged 99.94 in Tests, the next highest is Adam Voges, averaging 61.87.

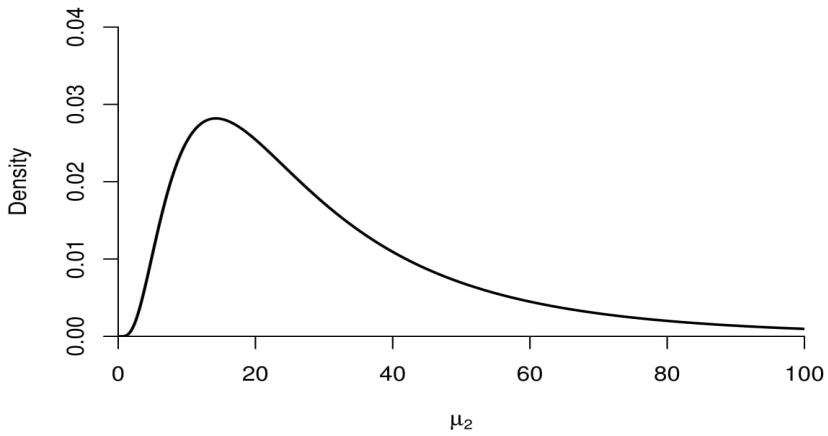


Figure 2.2. Prior probability density function for μ_2.

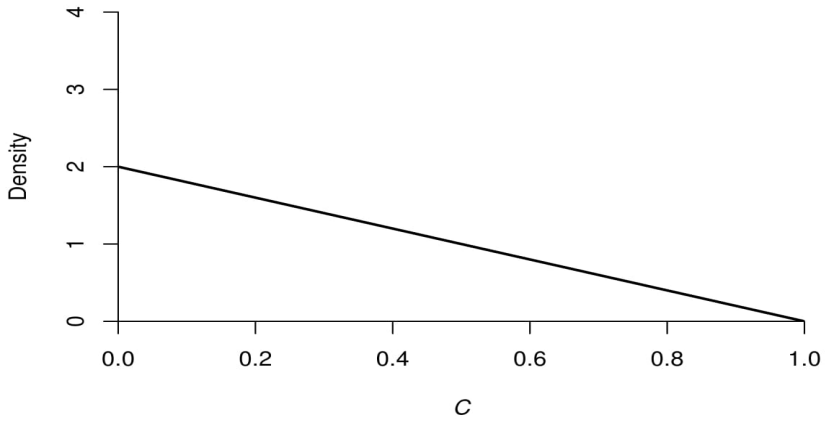


Figure 2.3. Prior probability density function for C.

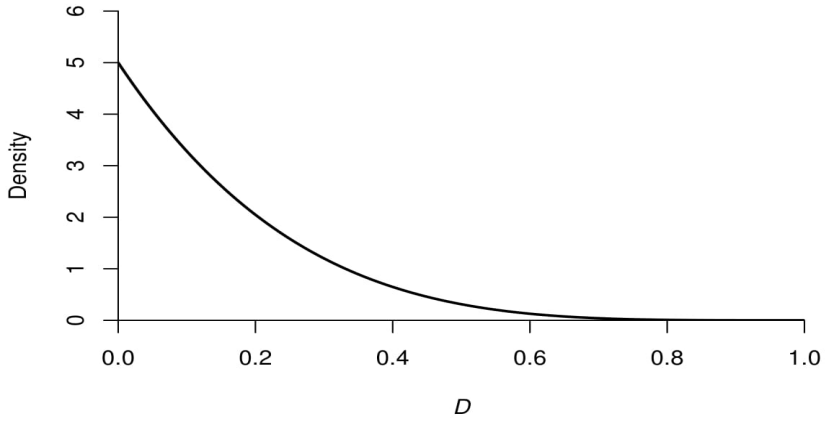


Figure 2.4. Prior probability density function for D.

The joint posterior distribution for μ_2, C and D is proportional to the prior times the likelihood function. We can then sample from the joint posterior distributions to make inferences about an individual player's initial batting ability (μ_1), 'eye-in' batting ability (μ_2) and the abruptness of the transition between these states (L).

Implementing the exponential varying-hazard model

To perform the computation, we used a Julia (Bezanson et al., 2014) implementation[2] of the nested sampling algorithm (Skilling, 2006) that uses Metropolis-Hastings updates (see Section 1.2.1). This allows us to easily obtain and sample from the posterior distributions of parameters μ_1, μ_2 and L, for each player, as well as computing the marginal likelihood.

For each player, we used 1000 nested sampling particles and 1000 MCMC steps per nested sampling iteration. As the model only contains three parameters, simpler MCMC schemes (or even simple Monte Carlo or importance sampling) would work here. However, we used nested sampling from the beginning as it allowed us to continue using the same method, even as our models increase in complexity (see Chapter 4), and carry out model selection trivially (Chapter 5).

2.2.3 Data

The data used were the career batting records of the players considered in the study and were obtained from Statsguru, the cricket statistics database on the Cricinfo website[3]. This was achieved using web scraping techniques with help of the R package `cricketr` (Ganesh, 2016). A range of variables are available for each innings, however we are primarily interested in the number of runs scored and whether or not the batsman was dismissed (for example see Table 2.1).

Test match data were chosen in favour of other formats, as the model assumptions are more likely to be sufficiently realistic. Players have more time to bat in Test matches and therefore scores are more likely to reflect a player's true batting

[2]https://github.com/eggplantbren/NestedSampling.jl
[3]http://www.espncricinfo.com/

Table 2.1: Example of a Test match batting career data file, including the number of runs scored, minutes batted, balls faced and mode of dismissal for each innings.

Runs	Mins	BF	4s	6s	SR	Pos	Dismissal	Inns	Opposition	Ground	Start-Date
46	158	103	6	0	44.66	1	Bowled	1	Pakistan	Christchurch	15-Mar-2001
73*	281	219	10	0	33.33	3	Not out	3	Pakistan	Christchurch	15-Mar-2001
106	422	280	14	1	37.85	1	Caught	2	Pakistan	Hamilton	27-Mar-2001
26	82	61	4	0	42.62	1	LBW	2	Australia	Brisbane	8-Nov-2001
57	93	69	6	0	82.60	1	LBW	4	Australia	Brisbane	8-Nov-2001

nature, rather than the specific match situation, such as batsmen tending to play more aggressively at the beginning and end of a team's innings in a one-day match.

As the varying-hazard model is an adaptation of the model specified in Brewer (2008), the same data set was used to assess model performance. This data set consists of an arbitrary mixture of retired batsmen, all-rounders and a bowler, each of whom enjoyed a long Test career during the 1990s and 2000s (see Table 2.2).

Table 2.2: Players analysed using the exponential varying-hazard model.

Player	Role	Country
C. Cairns	All-Rounder	New Zealand
N. Hussain	Batsman	England
G. Kirsten	Batsman	South Africa
J. Langer	Batsman	Australia
B. Lara	Batsman	West Indies
S. Pollock	All-Rounder	South Africa
S. Warne	Bowler	Australia
S. Waugh	Batsman	Australia

2.3 Results

2.3.1 Marginal posterior distributions

The model allows us to draw samples from the each of marginal posterior distributions for each parameter, for each player. To illustrate the practical implications of the results, posterior samples for former Australian captain Steve Waugh are shown in Figure 2.5.

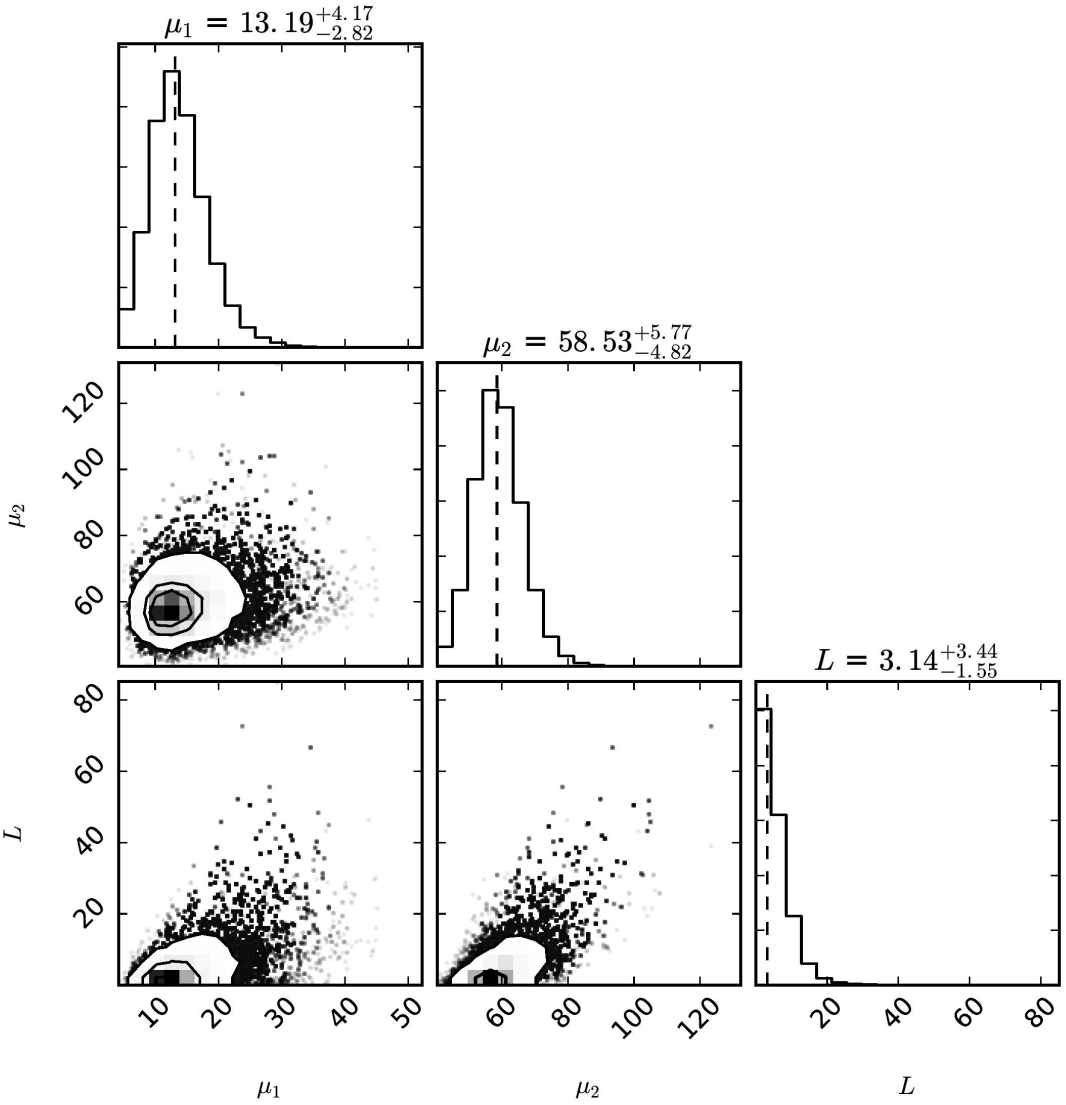


Figure 2.5. Posterior marginal distributions for μ_1, μ_2 and L for Steve Waugh. The contours represent the 50^{th}, 68^{th} and 95^{th} percentile limits. Created using the corner.py package (Foreman-Mackey, 2016).

The marginal distribution for μ_1 implies that Waugh arrives at the crease batting with the ability of a player with an average of 13.2 runs. After scoring about 3 runs, Waugh has transitioned approximately 63% of the way between his initial batting ability and 'eye-in' batting ability. Looking more closely at the effective average curve suggests Waugh reaches his 'eye-in' batting ability after scoring approximately 20 runs, at which point he bats like a player with an average of 58.5. Figure 2.6 gives a visual representation of these estimates.

The marginal distributions in Figure 2.5 are used to construct point estimates for the effective average curves (using $\mu(x; \mu_1, \mu_2, L)$ from Equation 2.9). These curves, seen in Figures 2.6 and 2.7, indicate how well individual players are batting given their current score, that is, the average number of runs they will score from a given score onwards.

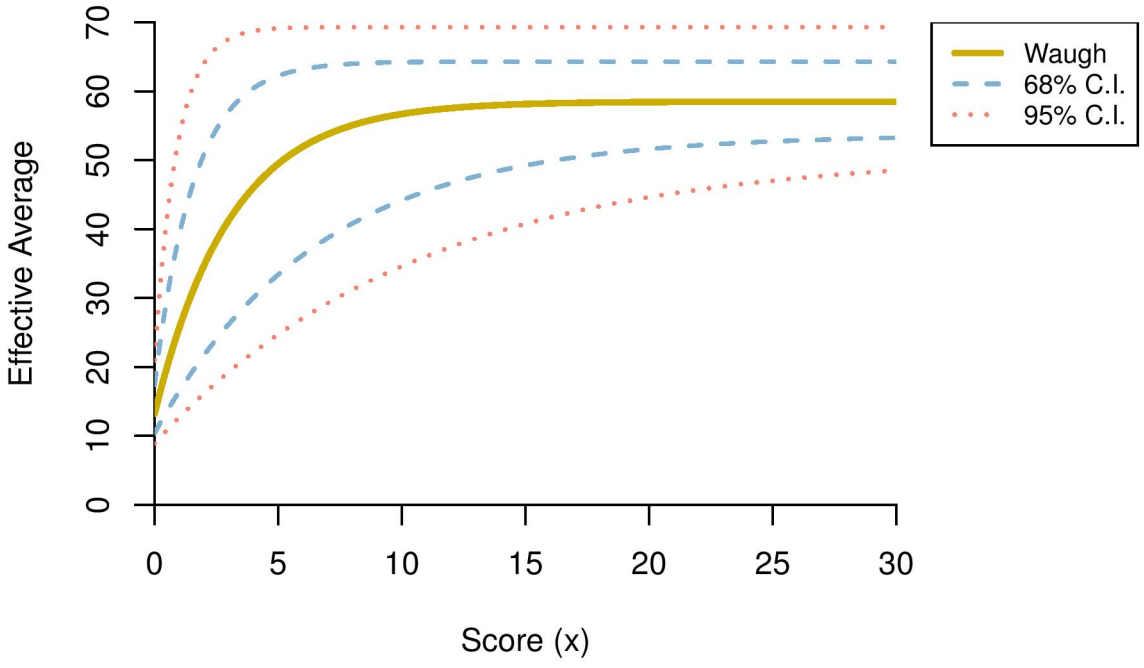


Figure 2.6. Plot of Steve Waugh's estimated effective average $\mu(x)$, illustrating how his batting ability changes with his current score. The blue and red lines represent 68% and 95% credible intervals.

2.3.2 Posterior summaries

Using the marginal posterior distributions, estimates and uncertainties were derived for the three parameters of interest for each player. The estimates take the form, posterior median ± standard deviation, and are presented in Table 2.4, together with each player's Test career record in Table 2.3. The median is used as the posterior distributions are not necessarily symmetric and some have relatively heavy tails.

Table 2.3: Test career batting records for analysed players.

Player	Matches	Innings	Not Outs	Runs	High-Score	Average	Strike Rate	100s	50s
C.Cairns (NZ)	62	104	5	3320	158	33.53	57.09	5	22
N.Hussain (ENG)	96	171	16	5764	207	37.18	40.38	14	33
G.Kirsten (SA)	101	176	15	7289	275	45.27	43.43	21	24
J.Langer (AUS)	105	182	12	7696	250	45.27	54.22	23	30
B.Lara (WI)	131	232	6	11953	400*	52.88	60.51	34	48
S.Pollock (SA)	108	156	39	3781	111	32.31	52.52	2	16
S.Warne (AUS)	145	199	17	3154	99	17.32	57.65	0	12
S.Waugh (AUS)	168	260	46	10927	200	51.06	48.64	32	50

Unsurprisingly, the players with the highest career averages (Brian Lara and Steve Waugh) appear to be the best players once they have their 'eye-in' (i.e. they have the highest μ_2 estimates). However, it is not necessarily these players who arrive at the crease batting with the highest ability. In fact, two of the players with the highest initial batting abilities, μ_1, are those with lower career Test averages, all-rounders Chris Cairns and Shaun Pollock. Interestingly, both players tend to bat in the middle to lower order and have lower estimates for μ_2, their 'eye-in' batting ability, suggesting they do not quite have the same batting potential as the other top order batsmen. This outcome may be due to initial batting conditions tending to be more difficult for batsmen in the top order, compared with those in the middle and lower order. Additionally, the result may derive from the aggressive nature in which Cairns and Pollock play, meaning even when they are dismissed early in their innings, they often return to the pavilion with some runs to their name. Analysing a larger sample of similar lower-order, aggressive batsmen would be useful for determining

Table 2.4: Parameter estimates and uncertainties for each analysed player using the exponential varying-hazard model. The logarithm of the marginal likelihood for the exponential varying-hazard model is presented alongside the logarithm of the Bayes factor, comparing the exponential varying-hazard and constant hazard models. 'Prior' indicates the prior point estimates and uncertainties.

Player	$\boldsymbol{\mu_1}$	**68% C.I.**	$\boldsymbol{\mu_2}$	**68% C.I.**	$\boldsymbol{L}$	**68% C.I.**	$\log_e(\mathbf{Z})$	$\log_e(\mathbf{Z/Z_0})$
C.Cairns	$16.6^{+6.4}_{-5.2}$	[11.4, 23.0]	$36.1^{+4.4}_{-3.8}$	[32.3, 40.5]	$2.3^{+4.6}_{-1.7}$	[0.6, 6.9]	−449.69	1.04
N.Hussain	$12.8^{+4.6}_{-3.2}$	[9.6, 17.4]	$40.8^{+4.3}_{-3.5}$	[37.3, 45.1]	$1.9^{+2.7}_{-1.2}$	[0.7, 4.6]	−714.52	5.88
G.Kirsten	$14.4^{+4.8}_{-3.4}$	[11.0, 19.2]	$53.9^{+6.6}_{-5.3}$	[48.6, 60.5]	$6.3^{+4.6}_{-2.9}$	[3.4, 10.9]	−769.79	10.00
J.Langer	$18.0^{+7.5}_{-4.8}$	[13.2, 25.5]	$49.2^{+5.0}_{-4.2}$	[45.0, 54.2]	$2.7^{+4.1}_{-1.8}$	[0.9, 6.8]	−800.83	3.95
B.Lara	$15.1^{+4.6}_{-3.5}$	[11.6, 19.7]	$61.8^{+5.7}_{-5.1}$	[56.7, 67.5]	$6.1^{+3.9}_{-2.7}$	[3.4, 10.0]	−1114.95	13.37
S.Pollock	$18.2^{+4.8}_{-4.3}$	[13.9, 23.0]	$37.4^{+5.8}_{-4.4}$	[33.0, 43.2]	$5.6^{+6.0}_{-3.5}$	[2.1, 11.6]	−526.15	1.83
S.Warne	$5.3^{+1.2}_{-0.9}$	[4.4, 6.5]	$21.2^{+2.1}_{-1.9}$	[19.3, 23.3]	$1.3^{+1.2}_{-0.8}$	[0.5, 2.5]	−679.77	15.60
S.Waugh	$13.2^{+4.2}_{-2.8}$	[10.4, 17.4]	$58.5^{+5.8}_{-4.8}$	[53.7, 64.3]	$3.1^{+3.4}_{-1.6}$	[1.5, 6.5]	−1032.36	13.98
Prior	$6.6^{+12.8}_{-5.0}$	[1.6, 19.4]	$25.0^{+27.7}_{-13.1}$	[11.9, 52.7]	$3.0^{+6.7}_{-2.3}$	[0.7, 9.7]	N/A	N/A

whether or not strike rate and batting position are in fact influential on a player's point estimate for parameter C (the size of μ_1 with respect to μ_2).

The marginal likelihood or evidence, was also computed for each player analysed using the individual player model. In this case we can use the evidence to compare the support for our varying-hazard model (Z), against a constant hazard model (Z_0) which has a Lognormal(log(20), 0.75^2) prior assigned to its constant effective average μ. The logarithm of the Bayes factor between these two models is included in Table 2.4 and suggests the varying-hazard model is favoured for all players. As the nested sampling method used is an MCMC process, these results are not exact, however the algorithm was run with a large number of particles and MCMC iterations and therefore the Monte-Carlo related errors are negligible.

These results are relatively consistent with Brewer (2008), who used a different model for $\mu(x)$; Cairns, Langer and Pollock are the best batsmen when first arriving at the crease, and Lara and Waugh have the highest 'eye-in' batting abilities. The actual point estimates were similar in most cases, though the present model has less uncertainty in values of μ_1 (most likely since $\mu_1 = \mu(0)$ in our model), but more uncertainty in μ_2 values. It is difficult to directly compare the transition variable L,

as Brewer (2008) used two parameters to capture the change between the two batting states.

2.3.3 Predictive hazard functions

The posterior summaries allow us to construct predictive hazard and predictive effective average functions, for a given player's next innings. This is a slightly different point estimate for $\mu(x)$ than the posterior mean or median.

Given a player's career batting data, $\boldsymbol{d}$, the predictive hazard function is obtained by calculating the posterior predictive distribution for a player's 'next' score given the data, and deriving the hazard function $H(x)$ corresponding to the predictive distribution using Equation 2.1. The predictive effective average function can then be derived from the predictive hazard function using Equation 2.9. For clarity, only the functions for four of the recognised batsman in the analysis were included (Gary Kirsten, Justin Langer, Brian Lara and Steve Waugh).

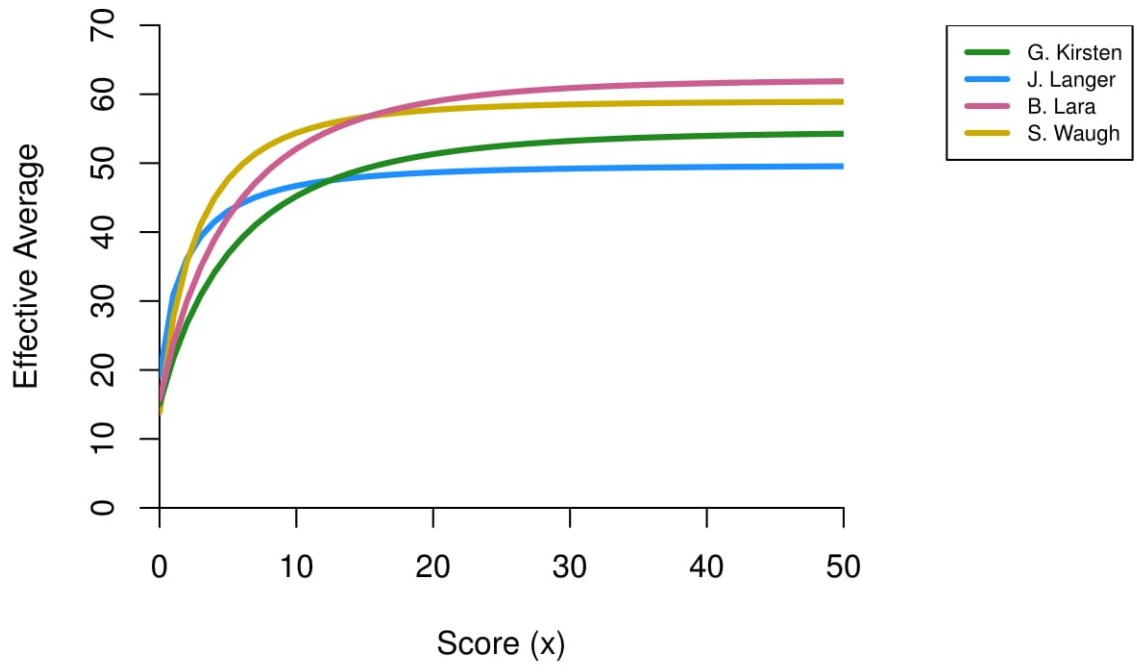


Figure 2.7. Predictive effective average functions, $\mu(x)$, for Kirsten, Langer, Lara and Waugh.

Figures 2.7 and 2.8 gives a visual representation of the posterior summaries in Table 2.4. Of the four players shown, Waugh has the lowest effective average when first arriving at the crease. However, Waugh gets his 'eye-in' relatively quickly and appears to be batting better than the others after scoring just 2 or 3 runs. Not until scoring approximately 15 runs does Lara overtake Waugh in terms of effective average, suggesting Lara is a better batsman when set at the crease (Equation 2.16), but takes longer to 'get his eye-in' (Equation 2.17).

$$P(\mu_{2\ Lara} > \mu_{2\ Waugh}|\boldsymbol{d}) = 0.66 \tag{2.16}$$

$$P(L_{Lara} > L_{Waugh}|\boldsymbol{d}) = 0.75 \tag{2.17}$$

An interesting comparison can also be made between Kirsten and Langer, two opening batsmen with identical[4] career Test batting averages of 47.27. Langer arrives at the crease with a higher initial batting ability than Kirsten (Equation 2.18) and is

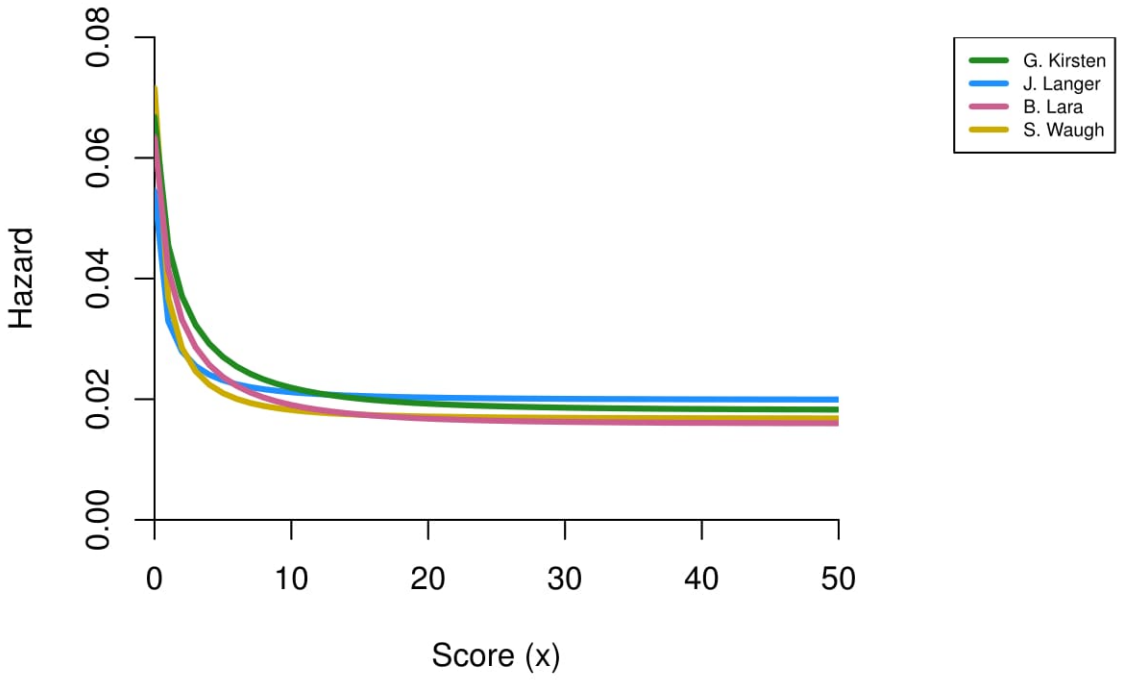


Figure 2.8. Predictive hazard functions, $H(x)$, for Kirsten, Langer, Lara and Waugh.

[4]Identical within two decimal places.

also quicker to get his 'eye-in' (Equation 2.19).

$$P(\mu_{1\ Langer} > \mu_{1\ Kirsten}|\boldsymbol{d}) = 0.70 \tag{2.18}$$

$$P(L_{Langer} < L_{Kirsten}|\boldsymbol{d}) = 0.77 \tag{2.19}$$

Only after scoring about 13 runs, does Kirsten look to be playing better than Langer in terms of batting ability. This arguably makes Langer the preferred choice for an opening batsmen as it suggests he is less susceptible at the beginning of his innings and is more likely to succeed in his job as an opener, seeing off the new ball and opening bowlers.

2.4 Limitations and conclusions

The effective average curves derived from the exponential varying-hazard model in Section 2.3.3 are useful for identifying potential strengths and weaknesses during a batsman's innings. However, it is worth noting the model excludes parameters that are very important within a cricketing context. While all batsman will admit scoring runs often results in a boost in confidence, simply facing a delivery (and not scoring from it) will help a batsman become more used to the pace and bounce of the pitch, and in turn aid in the process of getting their 'eye-in'. Some batsmen may even benefit from standing at the non-striker's end and watching their partner face several deliveries.

Therefore, in order to come to more substantive conclusions as to how well a player is batting at a given stage of their innings, it is worthwhile considering the variables 'balls faced' or 'minutes batting', in conjunction with runs scored. Such a model may be considered in the scope of future work, however, incorporating these additional variables into an accurate working model is no easy feat, as it introduces all sorts of complex interactions between variables.

In our estimation of the model parameters, we have assumed a batsman's ability is constant over the course of their career. In reality, it is far more realistic that some

time dependent effects exist between parameters μ_1, μ_2 and L. Temporal variation may exist on two scales, long-term changes due to factors such as age and experience, and short-term changes due to opposition, player form and confidence. Allowing the model parameters to vary across multiple innings may give us the ability to answer more difficult questions, such as how long it takes a new Test batsmen to find their feet on the international stage and start performing at their best. A player where this would be especially applicable would be former New Zealand captain, Daniel Vettori, who spent the first six years of his career (1997-2002) batting at numbers 9, 10 and 11, averaging just 16.26. However, over the remainder of his Test career (2003-2014) he averaged 36.47, and grew into a valuable all-rounder, frequently batting at numbers 6, 7 and 8.

Furthermore, temporal effects may also exist *during* a batsman's innings. In imposing the restriction that the hazard function must be monotonically decreasing (and therefore that the effective average is monotonically increasing), our estimates for a player's ability are not as erratic as those in Kimber & Hansford (1993) and Cai et al. (2002). However, it is entirely plausible, if not probable, that effects between certain scores and a player's effective average exist. For example, it is not uncommon to see batsmen lose concentration after batting for a long period of time or, before or after passing a significant milestone (e.g. scoring 50, 100, 200).

Therefore, in Chapter 4 we introduce more flexible models, which allow for a batsman's effective average to fluctuate, even after getting their 'eye-in', rather than plateauing after a certain score.

Chapter 3

Hierarchical analysis of New Zealand opening batsmen

3.1 Overview

In this chapter, the exponential varying-hazard model is applied using a hierarchical model structure, allowing us to make generalised inferences regarding a wider group of players; in this case, opening batsman who have represented New Zealand in Test matches, at the international level.

The model structure is redefined in Section 3.2, such that the prior for μ_2 is defined conditional on hyperparameters ν and σ. Posterior inference for ν and σ allows us to quantify the typical batting abilities and transition speeds for New Zealand opening batsmen since 1990.

In Section 3.3.2, these results are used to make an informed prediction concerning the batting abilities of the next opening batsman to debut for New Zealand. Comparisons are made between New Zealand's best performing retired opening batsman, Mark Richardson, and current opening batsman Tom Latham, identifying areas of particular strength and those which require improvement.

3.2 Model structure

While knowing the performance of individual players is useful, we can generalise our inference to a wider group of players by implementing a hierarchical model structure. Instead of applying the prior $\mu_2 \sim \text{Lognormal}(\log(25), 0.75^2)$ to each player, we define hyperparameters $\eta = (\nu, \sigma)$ such that the prior for each player's μ_2 is

$$\mu_{2,i}|\nu, \sigma \sim \text{Lognormal}(\log(\nu), \sigma^2). \tag{3.1}$$

When we infer ν and σ from the data for a group of players, we can quantify the typical μ_2 value the players are clustered around using ν, while σ describes how much μ_2 varies from player to player.

To apply the hierarchical model, each player in the group of interest was analysed using the exponential varying-hazard model in Chapter 2. The results were then post-processed to reconstruct what the hierarchical model would have produced. This is a common technique for calculating the output of a hierarchical model without having to analyse the data for all players jointly. Hastings (1970) suggested using MCMC samples for this purpose.

3.2.1 Prior specification

The hierarchical model is implemented by writing the prior for μ_2 conditional on hyperparameters $\eta = (\nu, \sigma)$, as Lognormal($\log(\nu)$, σ^2), rather than using a common Lognormal($\log(25)$, 0.75^2) prior for all players. The idea is to gain an understanding of the posterior distributions for ν and σ, rather than μ_2 directly. Whereas informal prior knowledge of cricket was used to assign the original Lognormal($\log(25)$, 0.75^2) prior, the hierarchical model does this more explicitly, as the prior for a player's parameters is informed by the data from other players. The priors over parameters C and D were kept constant (recall $\mu_1 = C\mu_2$ and $L = D\mu_2$).

We assigned flat, uninformative, Uniform(1, 100) and Uniform(0, 10) priors for the hyperparameters $\eta = (\nu, \sigma)$ respectively. These priors are conservatively wide to

ensure we do not miss sampling any areas of high likelihood in the parameter space. The full model structure is therefore

$$\nu \sim \text{Uniform}(1, 100) \tag{3.2}$$

$$\sigma \sim \text{Uniform}(0, 10) \tag{3.3}$$

$$\mu_{2,i}|\nu, \sigma \sim \text{Lognormal}(\log(\nu), \sigma^2) \tag{3.4}$$

$$C_i \sim \text{Beta}(1, 2) \tag{3.5}$$

$$D_i \sim \text{Beta}(1, 5) \tag{3.6}$$

$$\text{log-likelihood} \sim \sum_i (\text{Equation 2.6}) \tag{3.7}$$

where subscript i denotes the i^{th} player in our sample.

The marginal posterior distribution for the hyperparameters, given all of the data, may be written in terms of expectations over the individual players' posterior distributions computed using the exponential varying-hazard model in Chapter 2 (see e.g., Brewer & Elliott, 2014),

$$p(\nu, \sigma|\{\boldsymbol{d_i}\}) \propto p(\nu, \sigma) \prod_{i=1}^{N} \mathbb{E}\left[\frac{f(\mu_{2,i}|\nu, \sigma)}{\pi(\mu_{2,i})}\right] \tag{3.8}$$

where $f(\mu_{2,i}|\nu, \sigma)$ is the Lognormal$(\log(\nu), \sigma^2)$ prior applied to μ_2 for the i^{th} player, and $\pi(\mu_{2,i})$ is the Lognormal$(\log(25), 0.75^2)$ prior that was originally used to calculate the posterior for the i^{th} player. The expectation term (i.e., each term inside the product) can be approximated by averaging over the posterior samples for that player.

3.2.2 Data

The data used with the hierarchical model again come from Statsguru on the Cricinfo website. As the exponential varying-hazard model is able to pinpoint batsmen who are susceptible at the beginning of their innings, the hierarchical model was applied to opening batsmen.

In most conditions, the first overs of a team's innings are considered the most

difficult to face, as the ball is new and batsmen are not yet used to the pitch and weather conditions. In order to counter these difficult batting conditions, it may be expected that opening batsmen begin their innings batting closer to their peak ability than a middle or lower-order batsman. For the purposes of this discussion, in order to counter these difficult batting conditions, we assume that an ideal opening batsmen should be more 'robust', in the sense that the proportional difference between their initial and 'eye-in' batting abilities is smaller than most players (i.e. a high value for parameter C). This is by no means a global restriction imposed on opening batsmen around the world; in the past there have undoubtedly been great opening batsmen who were known to be vulnerable early in their innings. However as the exponential varying-hazard model gives us the ability to quantify initial and 'eye-in' abilities, it seems a reasonable assumption in the present context.

Given the author's nationality, the hierarchical model was applied to make generalised inference about opening batsmen who have represented New Zealand. As opening has been a position of concern for New Zealand for some time, all opening batsmen to have played for the national side since 1990 were included in the study. Any player who opened the batting for New Zealand for at least half of their Test innings during this time period was deemed an opening batsman and included in the hierarchical analysis[1]. The career records for the opening batsmen included in the analysis are presented in Tables A.1 and A.2 in Appendix A.

All innings for each batsman were included in the data set, even those that were not spent opening the batting. This was done under the assumption that players who are selected to open the batting are done so as they inherently possess batting traits and characteristics that coaches perceive to be desirable for opening. This assumes that a player who is a regular opening batsman, but has also spent some time batting at other top order batting positions, is unlikely to bat drastically differently between the two batting positions. The data set is up to date as of the completion of the 2017 home Test series versus Bangladesh (24^{th} January 2017)[2].

[1] This includes currently active players, Martin Guptill, Hamish Rutherford, Tom Latham and Jeet Raval.

[2] The next scheduled Test match for New Zealand is on 8^{th} March 2017.

3.3 Results

3.3.1 Hyperparameter summaries

New Zealand opening batsmen were analysed separately using the exponential varying-hazard model from Chapter 2. Posterior summaries were generated for each player and are presented in order of Test debut in Tables A.3 and A.4 (see Appendix A). Due to several players appearing in just a handful of matches, some uncertainties are fairly large.

Combining the posterior samples for each player and applying a Metropolis-Hastings algorithm, allows us to make posterior inferences regarding hyperparameters ν and σ, using the result from Equation 3.8. As the hyperpriors for both ν and σ were very conservative, the Metropolis-Hastings algorithm was run for a large number of iterations (one million). This allowed plenty of time for mixing and sampling from the joint distribution for both ν and σ. Both chains have a fairly low autocorrelation given the large number of iterations, and appear to have converged sufficiently as indicated by the traceplots for ν and σ, presented in Figures 3.1 and 3.2. As our starting point appears to be fairly typical of the joint posterior distribution, no burn-in period was necessary (Meyn & Tweedie, 1993).

The joint posterior distribution for ν and σ is shown in Figure 3.3 and represents just a small proportion of the area covered by the Uniform prior distributions, suggesting the data contained a lot of information about the hyperparameters.

The marginal posterior distribution for ν is shown in Figure 3.4, with the posterior predictive distribution for μ_2 overlaid. Our inference regarding ν can be summarised as: $\nu = 29.12^{+3.32}_{-3.14}$, while σ can be summarised as: $\sigma = 0.52^{+0.10}_{-0.07}$. These results suggest our subjectively assigned Lognormal$(\log(25), 0.75^2)$ prior in Chapter 2 was reasonably close to the actual frequency distribution of μ_2 values among this subset of Test cricketers.

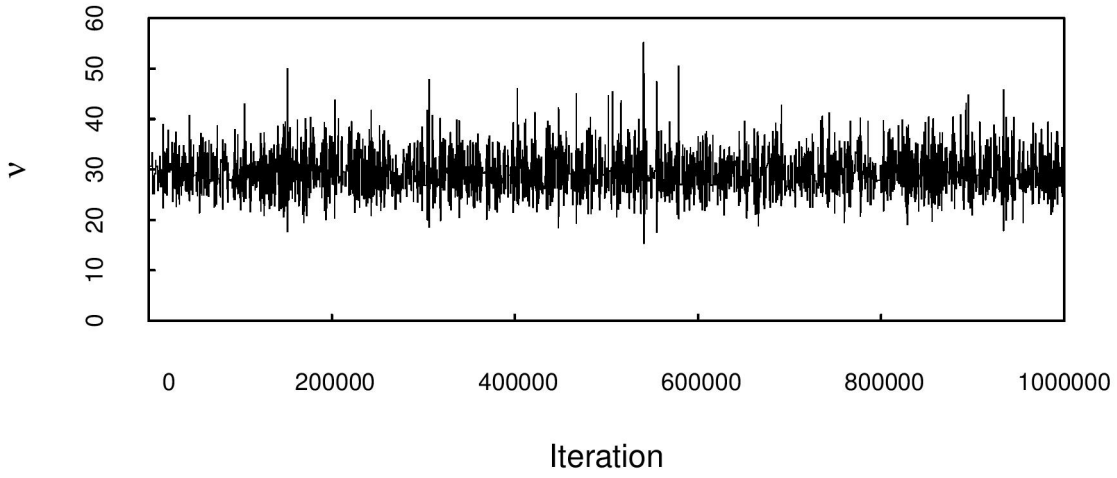


Figure 3.1. Traceplot for ν.

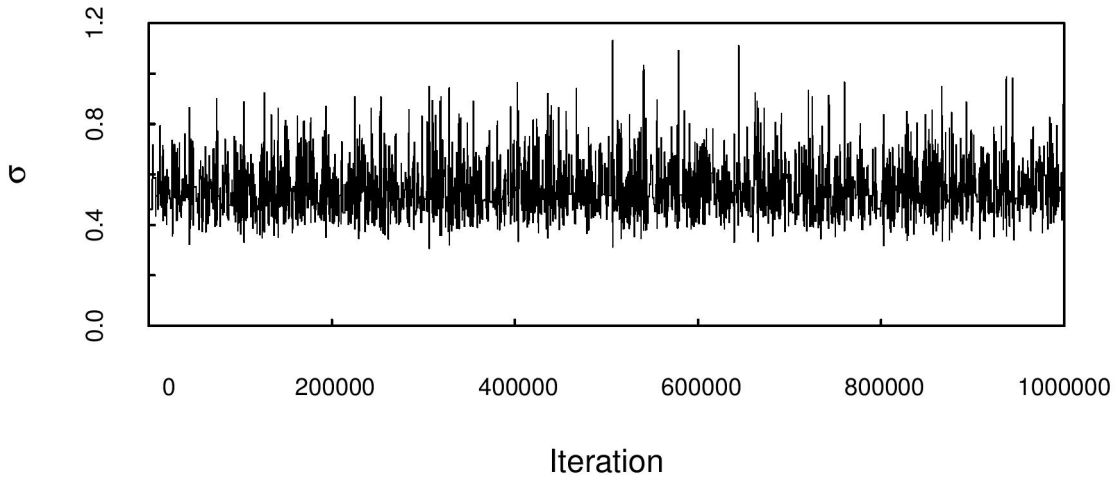


Figure 3.2. Traceplot for σ.

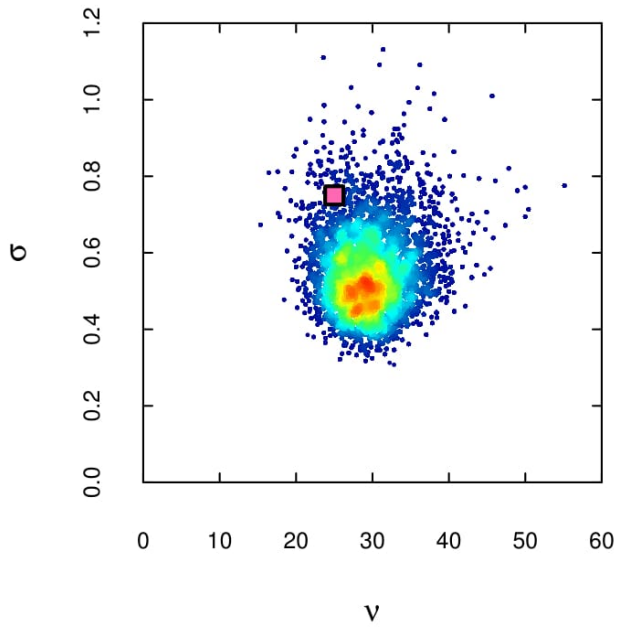


Figure 3.3. Joint posterior distribution for ν and σ describing the distribution of μ_2 values among the sample of New Zealand opening batsmen. The pink square indicates the position of the Lognormal(log(25), 0.75^2) prior.

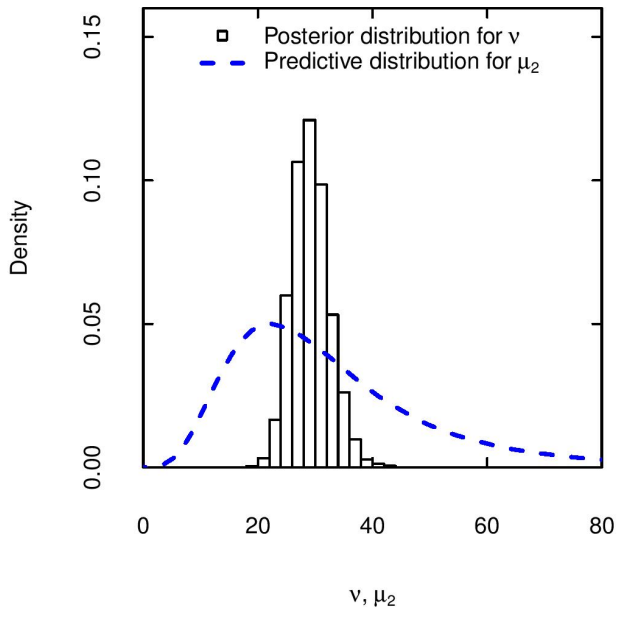


Figure 3.4. Marginal posterior distribution for ν with the predictive distribution for μ_2 overlaid.

3.3.2 Prediction for the next New Zealand opening batsman

Given the data, $\boldsymbol{d}$, and hyperparameters $\eta = (\nu, \sigma)$, we are able to make an informed prediction regarding the batting abilities, $\theta' = (\mu_1, \mu_2, L)$, of the 'next' opening batman to debut for New Zealand, using the result of Equation 3.9

$$\begin{aligned} p(\eta, \theta'|\boldsymbol{d}) &\propto p(\eta)\ p(\theta'|\eta)\ p(d|\theta', \eta) \\ &= p(\eta) \prod_{i=1}^{N} f(\theta'|\eta)\ p(d|\theta') \\ \therefore p(\theta'|\boldsymbol{d}) &\propto \int p(\eta) \prod_{i=1}^{N} f(\theta'|\eta)\ p(\boldsymbol{d}|\theta')\ d\eta. \end{aligned} \tag{3.9}$$

where θ' is obtained by marginalising over the hyperparameters using MCMC.

Our predicted estimates assume that the next player to debut is fairly typical of past openers and can be summarised as: $\mu_1 = 10.1^{+11.7}_{-5.8}$, $\mu_2 = 29.1^{+20.6}_{-12.1}$ and $L = 3.2^{+5.9}_{-2.4}$, presented in Table 3.1. For comparison, individual parameter estimates and uncertainties for all New Zealand openers in the data set are contained in Tables A.3 and A.4 in Appendix A.

The opening batsmen included in this study accounted for 727 separate Test innings. Given this moderate sample size, the uncertainties are somewhat large, although with more data we would expect more precise inferences and predictions.

Of course, this prediction must be taken with a grain of salt, as the New Zealand cricketing landscape has changed drastically since the 1990s. The ever-increasing amount of money invested in the game allows modern-day players to focus on being full-time cricketers. The structure of the domestic cricket scene has also improved, including better player scouting and coaching, resulting in the best local talents being

Table 3.1: Predictions for the batting abilities of the next opening batsman to debut for New Zealand.

	$\boldsymbol{\mu_1}$	**68% C.I.**	$\boldsymbol{\mu_2}$	**68% C.I.**	$\boldsymbol{L}$	**68% C.I.**
NZ Opener	$10.1^{+11.7}_{-5.8}$	[4.3, 21.8]	$29.1^{+20.6}_{-12.1}$	[17.0, 49.7]	$3.2^{+5.9}_{-2.4}$	[0.8, 9.1]

nurtured from a young age. Nevertheless, the prediction does highlight the difficulties New Zealand has had in the opening position. Few batsman with an 'eye-in' average, let alone a career average, of 29.1, would make many international sides on batting ability alone.

Figure 3.5 depicts the point estimates on the $\mu_1 - \mu_2$ plane for all New Zealand openers analysed in the study. All players fall within the 68% and 95% credible intervals of the prediction for the next opener, with the exception of Mark Richardson. Unsurprisingly, this suggests almost all players analysed are typical of New Zealand opening batsmen.

Since his debut in 2001, Richardson has widely been considered New Zealand's only world-class Test opener to play in the current millennium. Figure 3.5 certainly

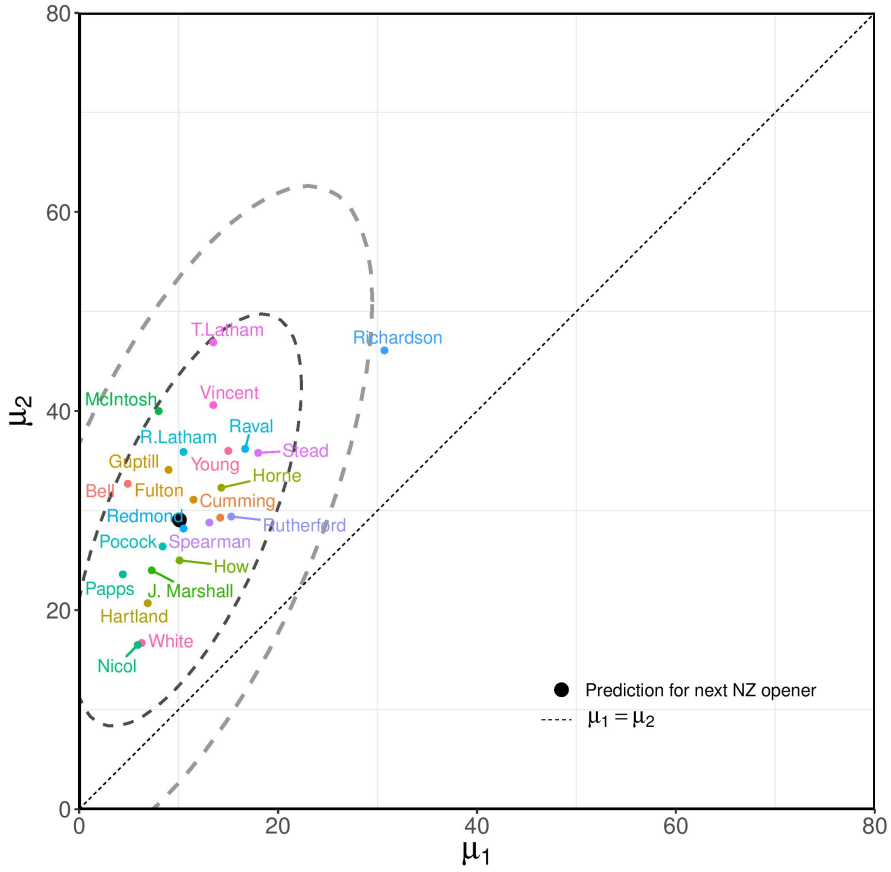


Figure 3.5. Point estimates for all analysed batsmen on the $\mu_1 - \mu_2$ plane. The prediction for the next New Zealand opening batsmen is represented by the black dot, including 68% (inner) and 95% (outer) credible intervals (dotted ellipses).

suggests Richardson is class apart from his compatriots, as he is the only player to fall outside the 68% and 95% credible intervals. Estimates for both Richardson's initial and 'eye-in' abilities are also considerably higher than the predicted abilities of the next opening batsman

$$P(\mu_{1\ \text{Richardson}} > \mu_{1\ \text{Predicted}}|\boldsymbol{d}) = 0.91$$
$$P(\mu_{2\ \text{Richardson}} > \mu_{2\ \text{Predicted}}|\boldsymbol{d}) = 0.81$$

In fact, Richardson's very high point estimate for $\mu_1 = 30.7$, suggests he begins his innings batting at least as well as the typical New Zealand opener, even once they have their 'eye-in', further highlighting the difficulty New Zealand has had with selecting capable opening batsman and/or Richardson's talent.

$$P(\mu_{1\ \text{Richardson}} \geq \mu_{2\ \text{Predicted}}|\boldsymbol{d}) = 0.51$$

Another player of interest is Tom Latham, whose batting abilities lie on the edge of the 68% credible interval on the μ_1 – μ_2 plane. In the last few years, Latham has established himself as New Zealand's current premier opening batsman in Test matches, reflected by the high estimate for his 'eye-in' batting ability.

Both Richardson and Latham have very similar 'eye-in' abilities (as seen in Table 3.2), suggesting Latham certainly has the talent to become one of New Zealand's stand-out modern-day openers.

$$P(\mu_{2\ \text{Richardson}} > \mu_{2\ \text{T.Latham}}|\boldsymbol{d}) = 0.47$$

However, it is Richardson's very high initial batting ability that sets him apart from other opening batsmen in the data set. Revisiting our assumption that an ideal opening batsman should have a high initial batting ability in order to counter the difficulties of facing the new ball, then Richardson edges Latham as the superior

Table 3.2: Parameter estimates and uncertainties for Mark Richardson and Tom Latham using the exponential varying-hazard model.

Player	$\boldsymbol{\mu_1}$	**68% C.I.**	$\boldsymbol{\mu_2}$	**68% C.I.**	$\boldsymbol{L}$	**68% C.I.**
M. Richardson	$30.7^{+8.5}_{-8.8}$	[21.9, 39.2]	$46.1^{+6.8}_{-5.6}$	[40.5, 52.9]	$3.6^{+6.9}_{-2.8}$	[0.8, 10.5]
T. Latham	$13.5^{+7.1}_{-4.7}$	[8.8, 20.6]	$46.9^{+8.9}_{-7.0}$	[39.9, 55.8]	$5.4^{+5.7}_{-3.5}$	[1.9, 11.1]

opener given his very high initial effective average:

$$P(\mu_{1\ \text{Richardson}} > \mu_{1\ \text{T.Latham}}|\boldsymbol{d}) = 0.93$$

This also suggests that Richardson is a very 'robust' batsman, in the sense that his initial batting ability is closer to his 'eye-in' batting ability than most other New Zealand openers. Latham on the other hand appears to be somewhat vulnerable early in his innings, compared with how well he bats once he has his 'eye-in', as indicated by the estimates for parameter C in Figure 3.7.

$$P(C_{\text{Richardson}} > C_{\text{Predicted}}|\boldsymbol{d}) = 0.83$$
$$P(C_{\text{T.Latham}} > C_{\text{Predicted}}|\boldsymbol{d}) = 0.39$$

Therefore, working on improving Latham's early innings batting may be the key for New Zealand cricket as he continues to build on the promising start made to his career as an opening batsman.

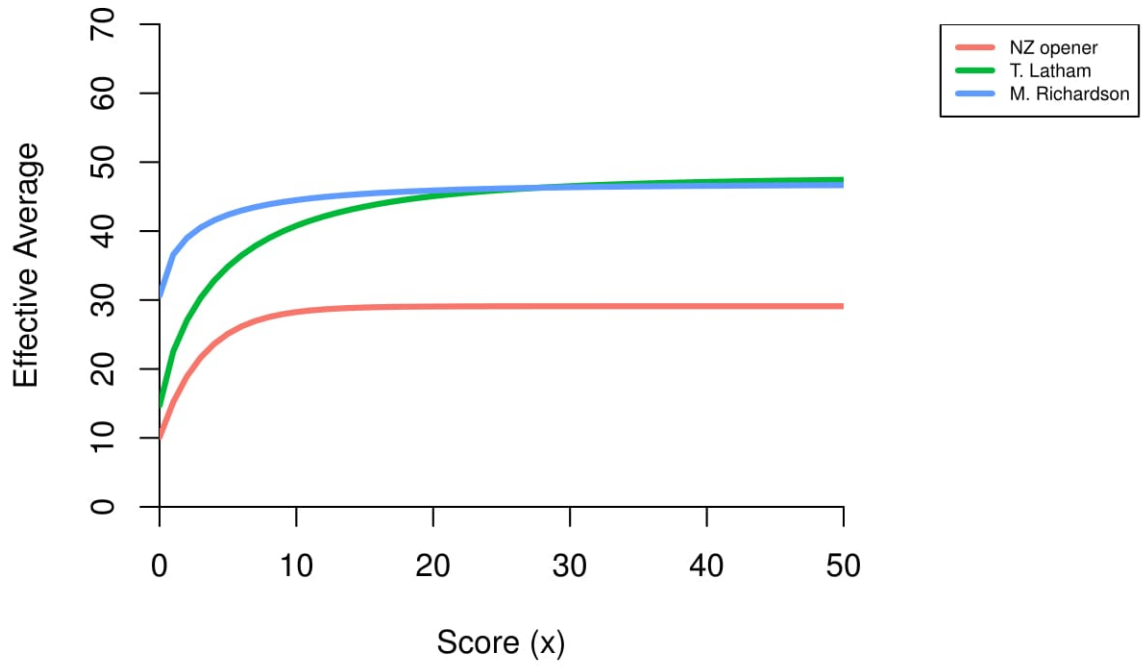


Figure 3.6. Predictive hazard functions in terms of effective average, $\mu(x)$, for Tom Latham, Mark Richardson and the next New Zealand opener.

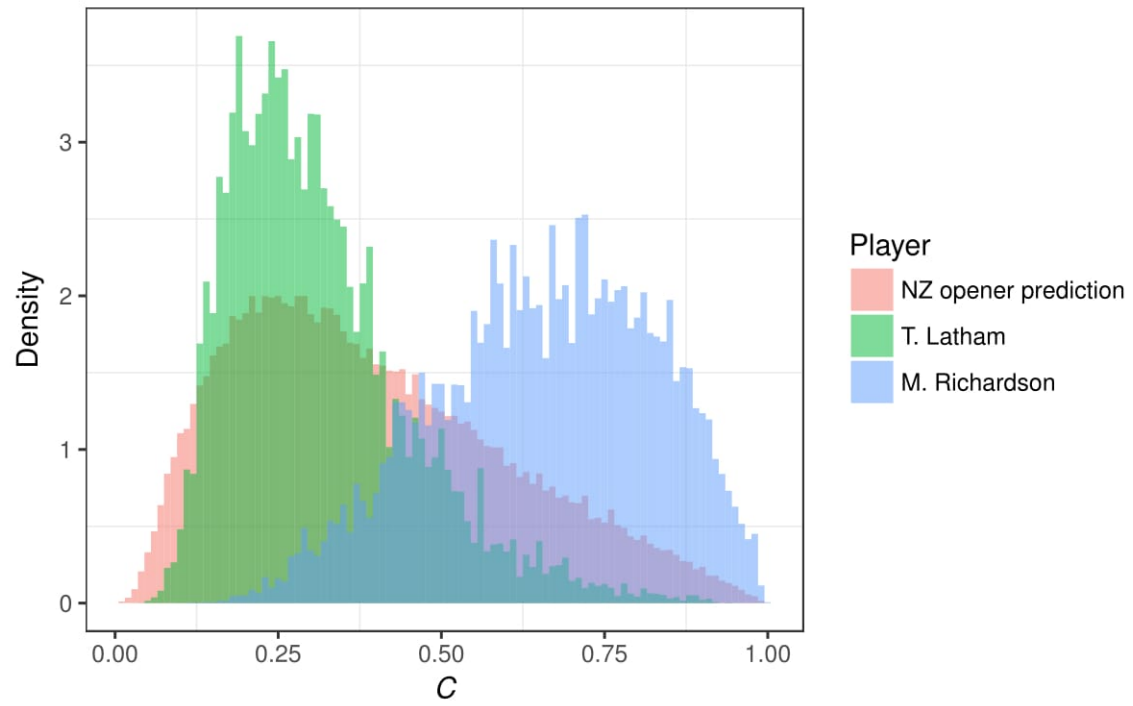


Figure 3.7. Histograms representing the marginal posterior distributions for parameter C for Tom Latham, Mark Richardson and the next New Zealand opener.

Chapter 4

Developing more flexible models

4.1 Overview

The exponential varying-hazard model detailed in Chapter 2 does a reasonable job at identifying batsmen who are especially capable or vulnerable early in their innings, compared with other players. However, as discussed in Section 2.4, the constraints imposed on this model also result in several limitations which are unrealistic within a cricketing context.

The main limitation of concern is that it is not realistic to believe that a batsman's ability monotonically increases over the course of their innings, before plateauing after a certain score. There are countless examples of batsmen taking a more cautious approach before passing significant milestones or becoming far more carefree once passing them. It is common to see opposition captains attempt to take advantage of batsman who are nearing these milestones, by setting more aggressive fields while the batsman may be thinking of their score, rather than giving their full attention to the ball being bowled at them.

This sort of behaviour certainly suggests there is likely some sort of temporal variation in ability between scores. Therefore to capture these changes in ability, our models must be afforded more flexibility in the parameterisation of the hazard, $H(x)$, and effective average, $\mu(x)$, functions. In Sections 4.2 and 4.3 two models are

outlined, each capturing temporal variation in batting ability during an innings in different ways.

The first model (Section 4.2) uses a Gaussian function to capture this variation by allowing for a temporary period of diminished or enhanced batting ability. This model uses fewer parameters but is also limited to only identifying a single period of variation in ability (hereafter referred to as the *Gaussian hazard model*).

The second model (Section 4.3) uses autoregressive terms to allow for more flexibility in the hazard function (hereafter referred to as the *AR(1) hazard model*). This model is much more flexible than the Gaussian hazard model, however requires far more parameters to be fitted, which can result in somewhat erratic hazard functions and excessively wide credible intervals.

Each model is applied to the same group of modern-day batsmen in Section 4.4 to determine whether or not any players exhibit significant score-related variation in ability, particularly when nearing scores of significance, such as during the 'nervous 90s'.

4.2 The Gaussian hazard model

A simple modification was made to the exponential varying-hazard model to allow for more flexibility in the effective average and hazard functions. Under the new model specification, the effective average function, $\mu(x)$, is able to exhibit temporary bumps or dips during a batsman's innings. This was achieved by multiplying the effective average function from the exponential varying-hazard model, $\mu(x; C, \mu_2, L)$, by the exponential of a Gaussian function. Such a model is not a drastic step away from the initial exponential varying-hazard model, but does allow for the identification of the strength and timing of any temporal variation in ability around particular scores, if it exists, for a given player.

The particular Gaussian functions under consideration contain three parameters: a strength parameter k, controlling the amplitude of the function, a width parameter ϕ, controlling the width of the function, and m, representing the midpoint of the func-

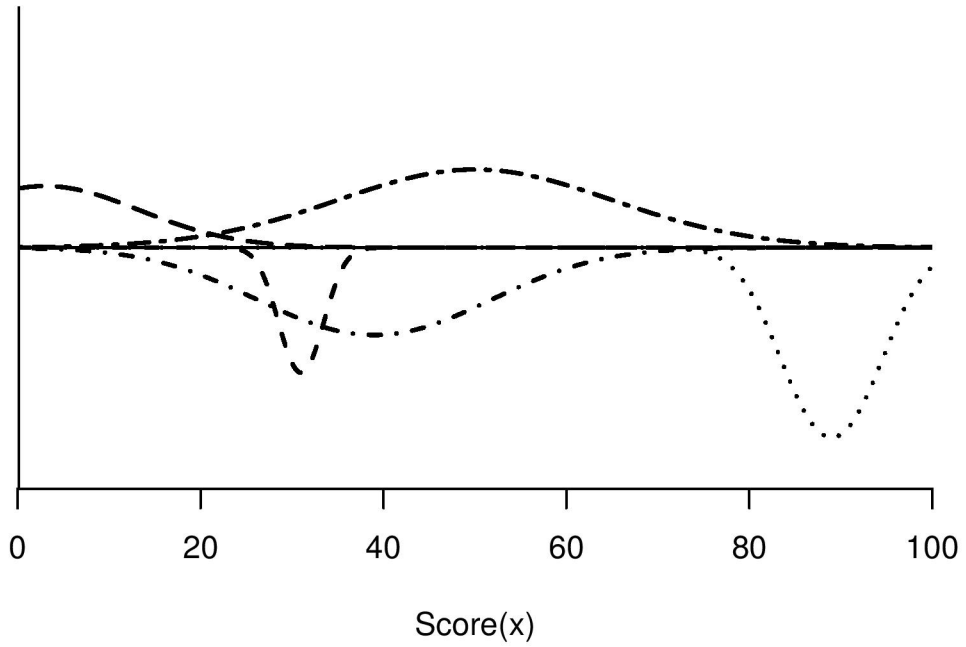


Figure 4.1. Examples of Gaussian functions with differing strengths, widths and midpoints.

tion (see Figure 4.1 for examples of plausible Gaussian functions). Mathematically, the Gaussian function can be written as

$$g(x; k, \phi, m) = -k \exp\left(\frac{-1}{2\phi^2}(x - m)^2\right) \tag{4.1}$$

The Gaussian function is often written with k, rather than $-k$ preceding the exponential. This makes no practical difference, the $-k$ simply indicates our default assumption is that players will tend to exhibit periods of decreased batting ability, rather than periods of increased ability.

4.2.1 Model structure

Parameterising the hazard function

The model likelihood follows the same derivation as that of the exponential varying-hazard model in Chapter 2. The only change required to implement the Gaussian

hazard function is to re-parameterise the effective average function to include the three new parameters (k, ϕ, m) used to produce the Gaussian element of the model. Like the exponential varying-hazard model, the hazard function takes the form

$$H(x) = \frac{1}{\mu(x) + 1}, \tag{4.2}$$

and relies on our parameterisation of the effective average function, $\mu(x)$.

If we define the effective average function from the exponential varying-hazard model, $\mu(x; C, \mu_2, L)$, as the 'underlying effective average', the Gaussian hazard model is obtained by multiplying the underlying effective average by the exponential of a Gaussian function. This gives the effective average for the Gaussian hazard model

$$\mu(x; C, \mu_2, D, k, \phi, m) = \left[\mu_2 + \mu_2(C-1)\exp\left(-\frac{x}{D\mu_2}\right)\right] \times \exp(g(x; k, \phi, m)). \tag{4.3}$$

To maintain positivity, the effective average function was modelled on the log-scale. Now, instead of multiplying the underlying effective average by the exponential of a Gaussian function, a Gaussian function from Equation 4.1 was added to $\log[\mu(x; C, \mu_2, L)]$

$$\log\left[\mu(x; C, \mu_2, D, k, \phi, m)\right] = \log\left[\mu_2 + \mu_2(C-1)\exp\left(-\frac{x}{D\mu_2}\right)\right] + g(x; k, \phi, m). \tag{4.4}$$

New effective average curves were produced, by back-transforming the effective average function by taking the exponential, which now allow for small deviations in batting ability during a batsman's innings. Examples of effective average functions allowed under the Gaussian hazard model can be seen in Figure 4.2, which combine the effective average functions from Figure 2.1 in Chapter 2, with the Gaussian functions from Figure 4.1.

Therefore, the Gaussian hazard model is able to account for any significant fluctuations in batting ability over the course of a player's innings. These deviations in ability can be large or small in terms of both amplitude and the range of scores

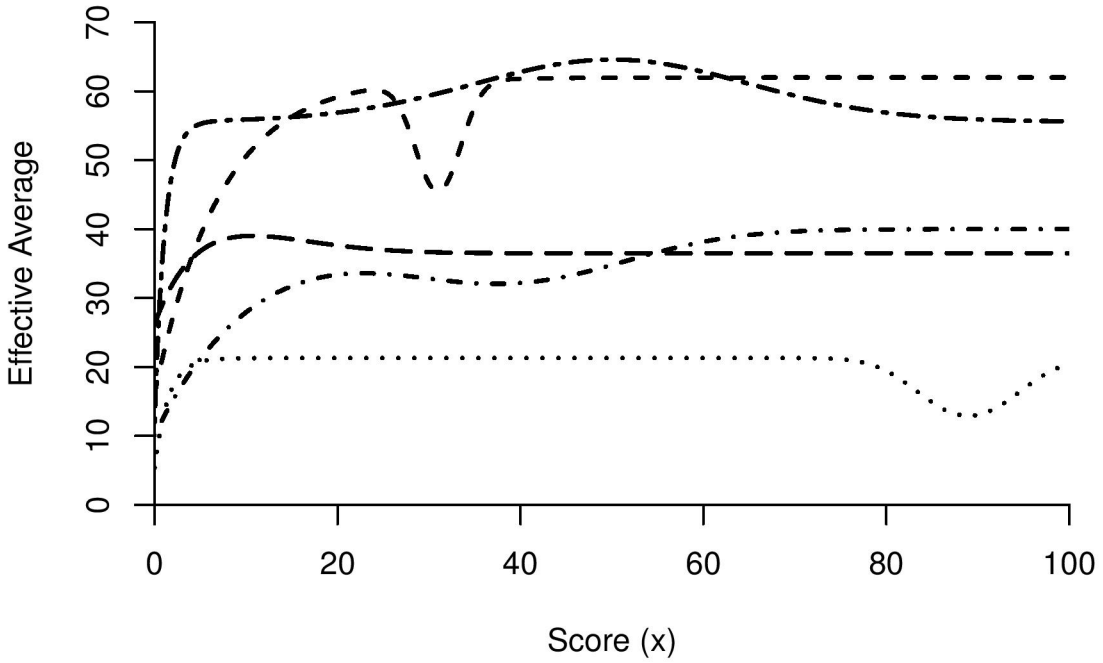


Figure 4.2. Examples of effective average functions, $\mu(x; C, \mu_2, D, k, \phi, m)$ allowed under the Gaussian hazard model, with varying levels and timings of temporal deviation in batting ability.

spanned. For example, a batsman who is notorious for being dismissed in the 'nervous 90s' would be expected to have an effective average function which exhibits a decline around scores of 90. This deviation in ability would present itself as a negative Gaussian function (i.e. a positive value for k) across the range of these scores.

Under this model, we still assume that the underlying effective average is a monotonically increasing function, therefore our constraint that a batsman's 'eye-in' effective average is larger than their initial effective average, $\mu_1 \leq \mu_2$, remains. Likewise, the assumption that the transition between the two batting states is no larger than the batsman's 'eye-in' effective average, $L \leq \mu_2$, is still enforced.

While the Gaussian function now allows the model to account for some temporal variation in batting ability, it is still unable to directly pick up on instances where batting may become harder due to a change in bowling or deterioration in local pitch and weather conditions. The model's real advantage lies in identifying the strength and timing of any systematic *score-based* deviations in batting ability, which may be

a result of a change in concentration levels due to a player's mood (Totterdell, 1999).

However, the model may indirectly pick up on bowling-related deviations in ability. For example, in almost every Test innings[1], the fielding side will opt to open the bowling with pace bowlers. Therefore, opening batsmen will tend to spend the majority of their early innings facing seam bowling, rather than spin. If the fielding side is unable to dismiss the opening batsmen cheaply, they may turn to spin once the batsmen have scored a moderate amount of runs, perhaps between 20 and 40. In this case, it may not be unrealistic to see an opening batsman exhibit a deviation in ability around scores in this range, which could hypothetically be a result of a change of bowling type.

Likewise, the model may also indirectly identify deviations in ability which are brought about by a change in tactics from the fielding team. As previously mentioned, there are plenty of examples of batsmen taking a more cautious approach before passing significant milestones, or becoming far more carefree once passing them. The model should be able to identify, (a) batsmen who frequently succumb to the pressure exerted by aggressive fields that are sometimes set to players nearing a significant score, and (b) batsmen who often give their wicket away once passing a significant score.

Prior specification

Once again, to formally specify the model, priors need to be assigned to each of the model parameters.

The priors for parameters C, μ_2 and D were chosen to remain the same as in the exponential varying-hazard model in Chapter 2. While it is likely that a player's hazard function exhibits temporal variation batting ability, it is difficult to directly translate our cricketing knowledge into informative priors for parameters k, ϕ and m. Therefore, relatively wide, conservative priors were assigned to each of these parameters, with each prior chosen to be independent of the priors on all other model

[1]With the possible exception of Test matches played on the spin-friendly, sub-continental pitches found in the likes of India, Sri Lanka and Bangladesh.

parameters.

To allow for reasonable fluctuations in ability in both directions (i.e. either an increase or decrease in batting ability at certain scores), a Uniform(-1, 1) prior was chosen for k. Likewise, to allow for both short and long periods of temporal variation in ability, a fairly wide Uniform(0, 20) prior was chosen for ϕ.

As the midpoint of the Gaussian function must coincide with a number of runs scored, m is restricted to the interval $[0, \infty)$. Since the highest individual score made in a Test is 400 not out[2], a very wide Uniform(0, 400) prior was assigned for m to allow for fluctuations in ability to occur at any stage during a batsman's innings.

Therefore, the overall Bayesian model specification for the Gaussian hazard model is

$$\mu_2 \sim \text{Lognormal}(\log(25), 0.75^2) \tag{4.5}$$

$$C \sim \text{Beta}(1, 2) \tag{4.6}$$

$$D \sim \text{Beta}(1, 5) \tag{4.7}$$

$$k \sim \text{Uniform}(-1, 1) \tag{4.8}$$

$$\phi \sim \text{Uniform}(0, 20) \tag{4.9}$$

$$m \sim \text{Uniform}(0, 400) \tag{4.10}$$

$$\text{log-likelihood} \sim \text{Equation (2.6)} \tag{4.11}$$

Implementing the Gaussian hazard model

The priors assigned to k, ϕ and m are most likely far too conservative; realistically there are no players with enough data at scores of 200+ to be confident of an increase or decrease in ability at such high scores. Very few players pass the milestone of a double century more than once, let alone at all during their entire career. Instead, we expect any meaningful inference to be made at scores between 0 and just past 100, where the majority of our data lie.

Like the exponential varying-hazard model, the nested sampling algorithm that

[2]Scored by Brian Lara for the West Indies vs. England in 2004.

uses Metropolis-Hastings updates (Skilling, 2006) was used for the Gaussian hazard model. However, a consequence of using such wide, uninformative priors is that the nested sampling algorithm must run for a larger number of iterations to effectively compress the parameter space beyond the posterior. Again, 1000 nested sampling particles were generated for each player, with 1000 MCMC steps per iteration.

While this model has three more parameters than the exponential varying-hazard model, the total number of parameters (6) is still low, meaning a sampling technique such as nested sampling is still overly complex for the Gaussian hazard model. However, we are not only aiming to determine whether a player exhibits temporal variation in batting ability, but also which model best captures this variation. Nested sampling allows us to easily compare these more flexible models using the marginal likelihoods, which is a primary result of the nested sampling algorithm.

4.3 The AR(1) hazard model

A second solution to affording the effective average and hazard functions increased flexibility, is to allow for a fluctuation in ability at each individual score. Such a model allows for more flexibility than the Gaussian hazard model, as temporal variation in ability is no longer restricted to a particular range of scores, bounded by the fitted Gaussian function.

This increased flexibility is achieved by estimating the underlying effective, average function, $\mu(x; C, \mu_2, D)$, in the same manner as the exponential varying-hazard model, as well as estimating a unique parameter, s_x, for each score, x, where s_x indicates the proportional deviation in batting ability at score x. For example, a value of $s_{55} = 2$, implies that a player bats with twice their underlying ability when on a score of 55. A batsman's effective average function can then be constructed by multiplying the underlying effective average function, $\mu(x)$, by each unique score parameter, s_x, as seen in Figure 4.3. Therefore, under the assumptions of this model, we would estimate batting abilities for scores 0 to 99, by using 100 independent score parameters, $\{s_x\} = \{s_0, ..., s_{99}\}$, as well as parameters C, μ_2 and D.

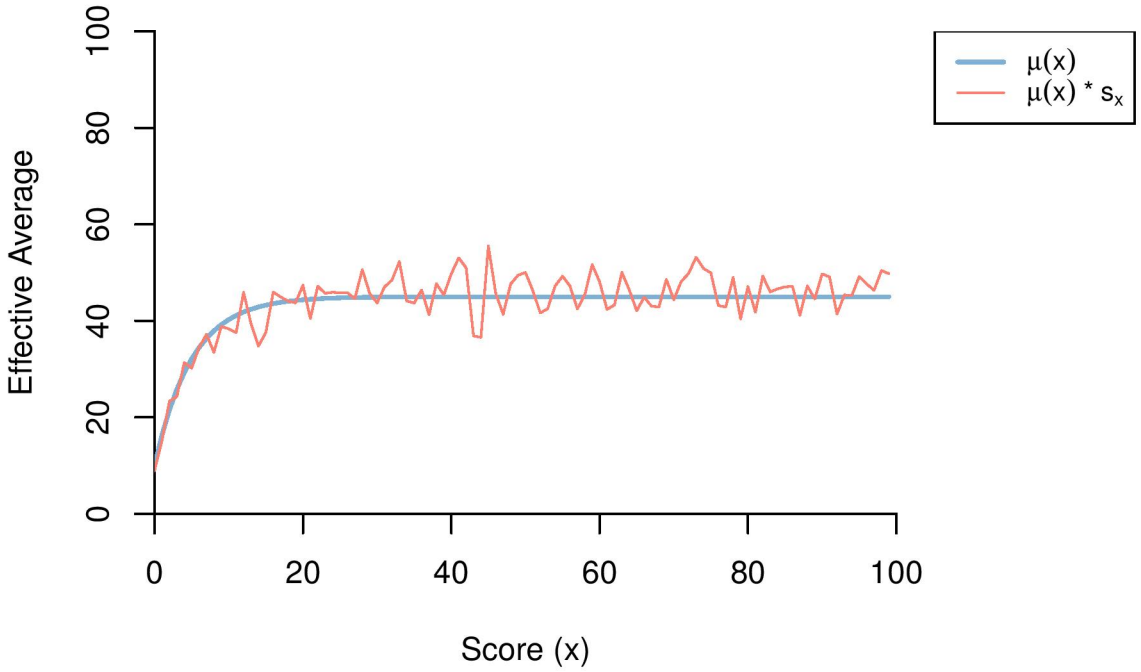


Figure 4.3. Example of an underlying effective average function, $\mu(x)$, which has also been multiplied by a unique s_x parameter at each score, producing an effective average that varies with score.

Clearly, this model requires a large number of parameters to be estimated, which is both a benefit and burden in terms of our estimation of batting ability. On one hand, it allows for far more fluctuation in ability than any other model we have developed. However, given the large number of parameters to be estimated, it can also lead to large uncertainties and unreasonable fluctuations in the estimation of the effective average and hazard functions.

Redefining the s_x terms

Similar to the limitations of Kimber & Hansford (1993), these uncertainties in ability will become increasingly large as the data become more scarce at higher scores. Additionally, as the set of unique score parameters, $\{s_x\}$, are virtually unrestricted, the resulting estimates for the effective average and hazard functions will take estimates that fit too conveniently to the data.

Over the course of a player's innings, it is not reasonable to believe that batting

ability will fluctuate so excessively from score to score. Obviously, how well a player is batting on a score of 20 is going to have some impact on how well they are batting on a score of 21 and 22, especially since runs are not always scored in singles. In fact, as runs are often scored in boundaries (in allotments of 4 or 6 runs), a batsman who is currently on 20 runs, is just one shot away from any score between 21 and 26. Therefore, it makes sense to impose a restriction on the model, such that there is some sort of distance-based correlation between the unique score parameters, s_x.

Information theory and maximum entropy tell us that where a constraint is imposed on a probability distribution, the model which is most similar to the original model that satisfies the new constraint, should be chosen (Jaynes, 1957; Caticha & Giffin, 2006). In this case, specifying s_x as an autoregressive (AR) process allows us to continue treating each score-based deviation in ability as a random process, but also restricts the s_x terms to be linearly dependent on previous terms (Sivia & Skilling, 2006).

Defining y_x as an AR(1) process (an autoregressive process with an order of one) gives

$$y_x = \lambda + \alpha(y_{x-1} - \lambda) + \beta n_x. \tag{4.12}$$

Under this definition, λ is the stationary mean of the process (a constant), α is the parameter which determines the temporal dependence between the s_x terms, n_x are white noise terms and β is the variance of these noise terms.

As s_x represents the deviation in a player's batting ability in terms of a proportion of their underlying ability at score x, all s_x terms should be restricted to the interval $[0, \infty)$. Therefore, the s_x terms were modelled as an AR(1) process on the logarithmic scale with the stationary mean λ, set equal to 0, giving

$$\log(s_x) = \alpha \log(s_{x-1}) + \beta n_x. \tag{4.13}$$

Under this definition, the AR(1) process has the following properties

$$\mathbb{E}\left[\log(s_x)\right] = 0, \qquad \text{Var}\left[\log(s_x)\right] = \frac{\beta^2}{1-\alpha^2}. \tag{4.14}$$

Now, instead of s_x being allowed to vary freely in the parameter space, each s_x term has some sort of conditional dependence with previous s_x terms, controlled by the parameter α, defined by

$$\alpha = \exp\left(\frac{-1}{\tau}\right), \tag{4.15}$$

where τ is the decay time, representing the number of future terms that are informed by a previous s_x term. Rearranging Equation 4.15 gives

$$\tau = \frac{-1}{\log(\alpha)}. \tag{4.16}$$

For example, $\alpha = 0.9$ gives a value of $\tau \approx 10$, implying the value of s_x at a particular score informs the values of terms s_{x+1}, s_{x+2}, ..., s_{x+9} and s_{x+10}. That is, a player's batting ability at a score of 20 will inform batting ability at scores 21, 22, ..., 29 and 30.

4.3.1 Model structure

Again, the model likelihood for the AR(1) hazard model is the same as that of the exponential varying-hazard model in Chapter 2, as is the underlying effective average function, $\mu(x; C, \mu 2, D)$. To fully define the AR(1) hazard model, the autoregressive terms (the s_x terms) need to be included in the full specification of the effective average function.

As s_x was modelled on the logarithmic scale, the full specification of the effective average function is

$$\mu(x; C, \mu_2, D, \{s_x\}) = \left[\mu_2 + \mu_2(C-1)\exp\left(-\frac{x}{D\mu_2}\right)\right] \times s_x \tag{4.17}$$

where $\{s_x\}$ comes from Equation 4.13.

Since the underlying effective average function, $\mu(x; C, \mu 2, D)$, remains unchanged from the exponential varying-hazard model, the constraints $\mu_1 \leq \mu_2$ and $L \leq \mu_2$ are still imposed. However, given the flexibility in the model afforded by the set of $\{s_x\}$ values, it is technically possible that instances may arise where the estimate for a player's effective average at a score 0 is actually higher than their estimated effective average at a score where they they are deemed to have their 'eye-in'. It would however, require a highly unrealistic data set for such a situation to arise.

Prior specification

Like the Gaussian hazard model, the priors for C, μ_2 and D were kept the same as in the exponential varying-hazard model. Priors for parameters α, β and n, which control the s_x terms, were chosen to be independent of priors assigned to all other model parameters.

In order to allow for extended periods of increased or decreased batting ability, while maintaining a steady underlying 'eye-in' effective average, the s_x terms were chosen to be a stationary process. Imposing the constraint $|\alpha| \leq 1$ ensures the AR(1) process will be stationary and will not exhibit rapidly oscillating values (Hamilton, 1994). Therefore, a Beta(4, 1) prior was assigned to α, emphasising values closer 1. This prior represents a mean value of $\alpha = 0.8$, implying a decay time $\tau = 4.48$, indicating temporal dependence in ability between scores within roughly 4 runs of each other.

An exponential prior with mean = 0.1 was assigned to β. This prior allows for the deviation in ability from score to score to change by up to approximately 10%, which seems reasonable in the context of the effective average function.

Finally, a Normal(0, 1) prior was assigned to the set of noise parameters, $\{n_x\}$. For a given particle in each nested sampling iteration, s_x terms are constructed using common values for α and β. What sets each s_x term apart is its specific noise parameter, n_x. Rather than using nested sampling to specifically evolve the s_x terms

(as the model which uses entirely independent s_x terms does), the AR(1) hazard model evolves the parameters used to construct the s_x terms, α, β and $\{n_x\}$. As α and β are constant for all s_x terms, any temporal variation in batting ability identified by the model, is being identified by the set of noise parameters, $\{n_x\}$.

Under these prior specifications, the AR(1) hazard model has a full Bayesian specification of

$$\mu_2 \sim \text{Lognormal}(\log(25), 0.75^2) \tag{4.18}$$

$$C \sim \text{Beta}(1, 2) \tag{4.19}$$

$$D \sim \text{Beta}(1, 5) \tag{4.20}$$

$$\alpha \sim \text{Beta}(4, 1) \tag{4.21}$$

$$\beta \sim \text{Exponential}(\text{mean} = 0.1) \tag{4.22}$$

$$\{n_x\} \overset{iid}{\sim} \text{Normal}(0, 1) \tag{4.23}$$

$$\text{log-likelihood} \sim \text{Equation (2.6)} \tag{4.24}$$

Implementing the AR(1) hazard model

For the analysis of individual players, the AR(1) hazard model was chosen to estimate 401 noise parameters (one for each score from 0 and 400, inclusive). Unlike the exponential varying-hazard and Gaussian hazard models, using classic nested sampling will be computationally inefficient and possibly inaccurate, given the large number of parameters to be estimated. As a result of the high dimensional parameter space, it is plausible that likelihood functions which exhibit multimodality exist, which can be a problem when using standard nested sampling methods. Therefore, to deal with any issues which may arise due to the large number of parameters in the AR(1) hazard model, a C++ implementation of the diffusive nested sampling algorithm (Brewer et al., 2011; Brewer & Foreman-Mackey, 2016) (see Section 1.2.1) which calls Julia to evaluate the likelihood function was used.

The diffusive nested sampling algorithm was run using 5 particles, with enough nested sampling iterations to adequately evolve each particle. Unlike the nested sam-

pling methods used for the exponential varying-hazard and Gaussian hazard models, using a larger number of particles is not advised, as there are over 400 model parameters to evolve within each nested sampling particle. Nested sampling levels were created every 10000 likelihood values above the current likelihood threshold, with a maximum of 100 levels created. All other model tuning parameters were kept constant as per the recommendations of Brewer & Foreman-Mackey (2016).

4.4 Results

4.4.1 Model testing

In order to establish confidence in our flexible models' abilities to accurately capture temporal changes in batting ability, a fake data set was generated to test each of the models. In particular, the data set contained scores from a 'player' who exhibited an extreme case of the 'nervous 90s' over the course of a long career.

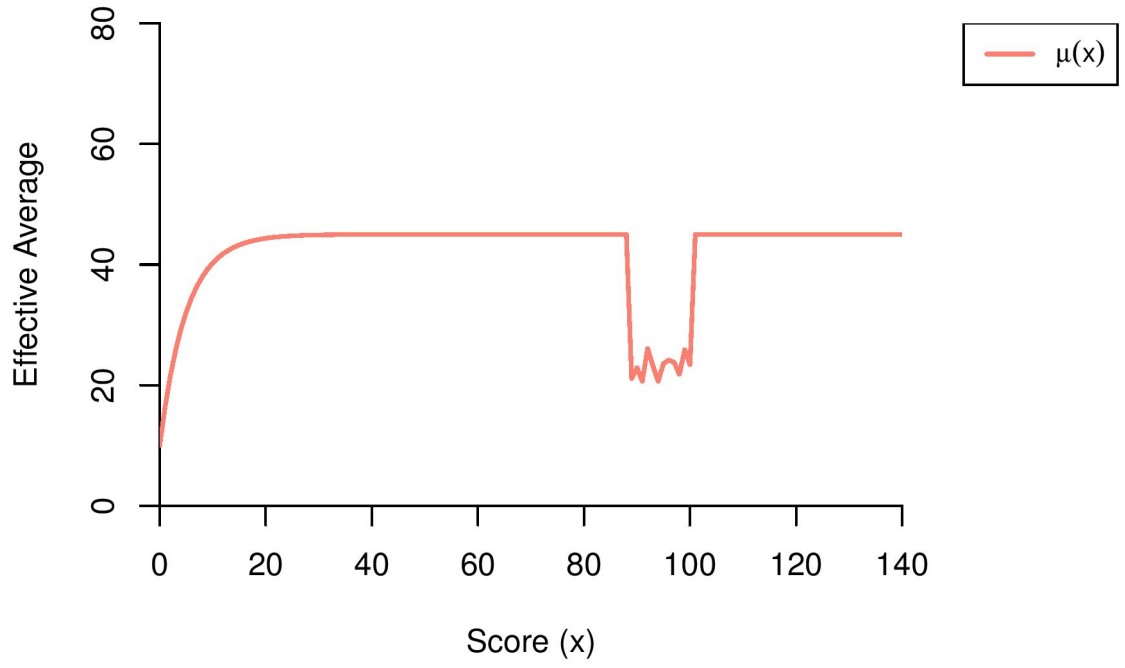


Figure 4.4. The underlying effective average function, $\mu(x)$, from which the fake data set was generated.

The synthetic data were generated by constructing a hazard function using arbitrary values for μ_1, μ_2 and L, in this case values of 10, 45 and 5 were chosen respectively. To ensure a disproportionate amount of scores in the 90s were present in the data set, a deviation in ability was imposed at scores 90 to 99. The resulting effective average function for the fake player can be seen in Figure 4.4. Scores were then simulated using the hazard function in Figure 4.5, with not out scores assigned randomly with a constant probability of 10%.

The resulting data set contained 1000 innings, which included 109 not out scores. The fake player averaged 39.6, with a top score of 245. Figure 4.6 depicts the frequency distribution of the scores in the data set, highlighting the difficulty the fake player had with getting through the 90s without being dismissed.

Using each of the Gaussian and AR(1) models, the data set was analysed to ensure the models are able to sufficiently identify the deviation in batting ability at scores in the 90s. For comparison, the data set was also analysed using the exponential varying-hazard model.

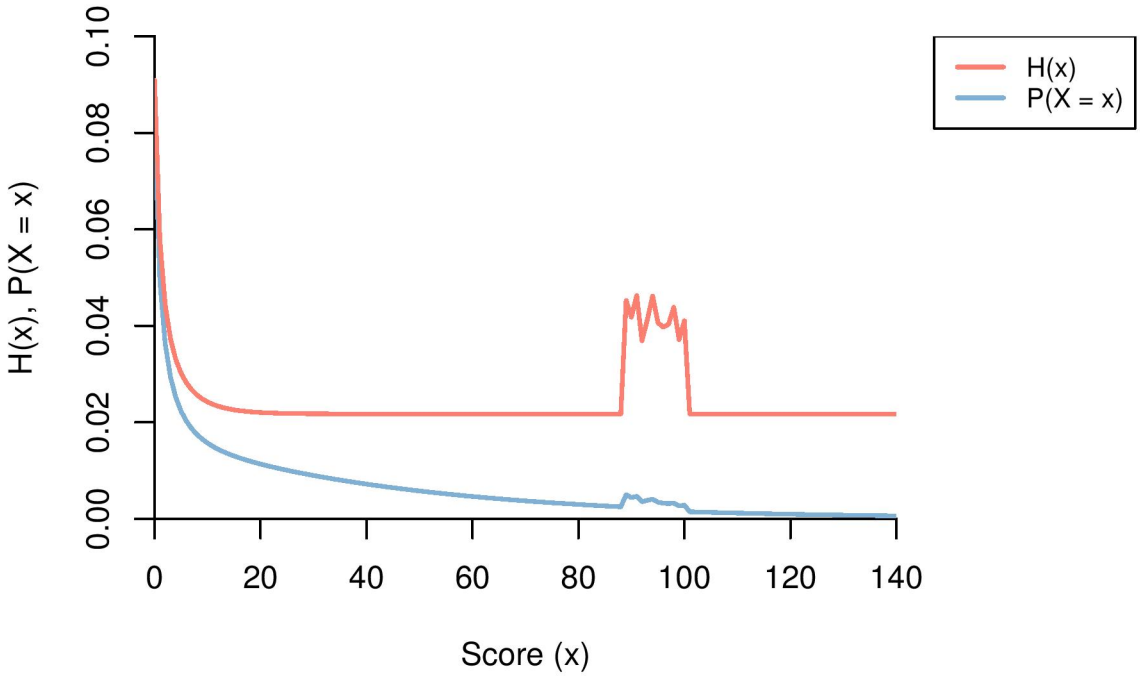


Figure 4.5. The underlying hazard and probability mass functions for the fake data set.

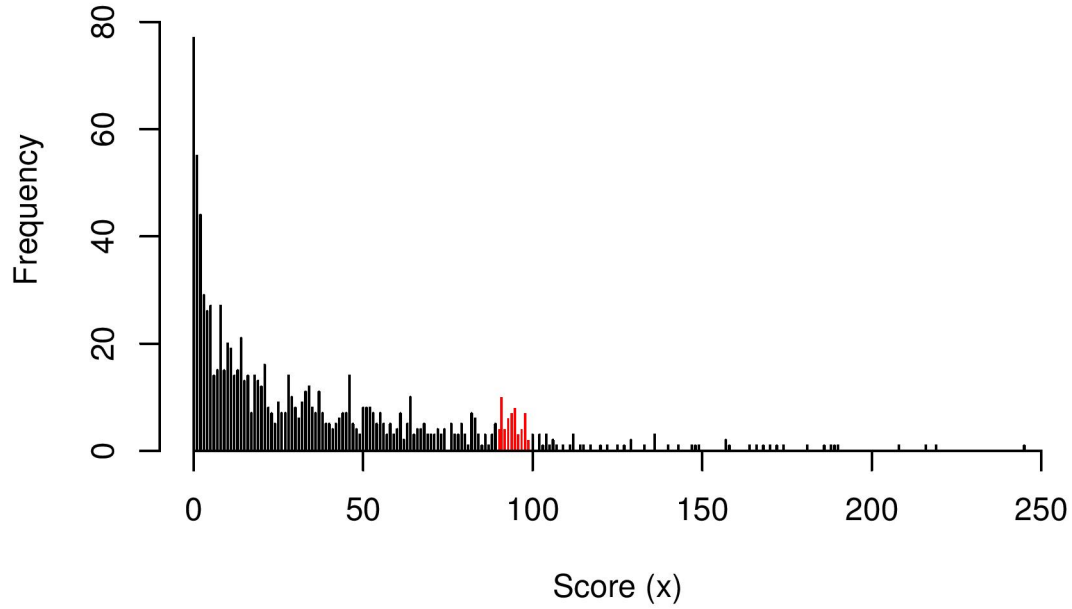


Figure 4.6. Histogram of scores from the simulated fake data set. Scores in the 90s are highlighted in red.

The Gaussian hazard model

The fake data set was analysed using the Gaussian hazard model, with samples drawn from the joint posterior distribution to obtain an estimate for the effective average function, for the fake player. Figure 4.7 suggests that the Gaussian hazard model did a reasonable job at capturing the temporal variation, indicating a deterioration in batting ability while on scores in and near the 90s.

The posterior parameter summaries for both the Gaussian hazard and exponential varying-hazard models are presented in in Table 4.1. As expected, the point estimate for m is during the 90s and the estimate for k is close to 1, indicating a large deviation

Table 4.1: Parameter point estimates for the fake data set for both the Gaussian hazard model and exponential varying-hazard model.

Model	$\boldsymbol{\mu_1}$	$\boldsymbol{\mu_2}$	$\boldsymbol{L}$	$\boldsymbol{k}$	$\boldsymbol{\phi}$	$\boldsymbol{m}$
Gaussian hazard	12.7	50.2	5.8	0.95	6.7	95.8
Exponential varying-hazard	12.6	46.9	4.9	-	-	-

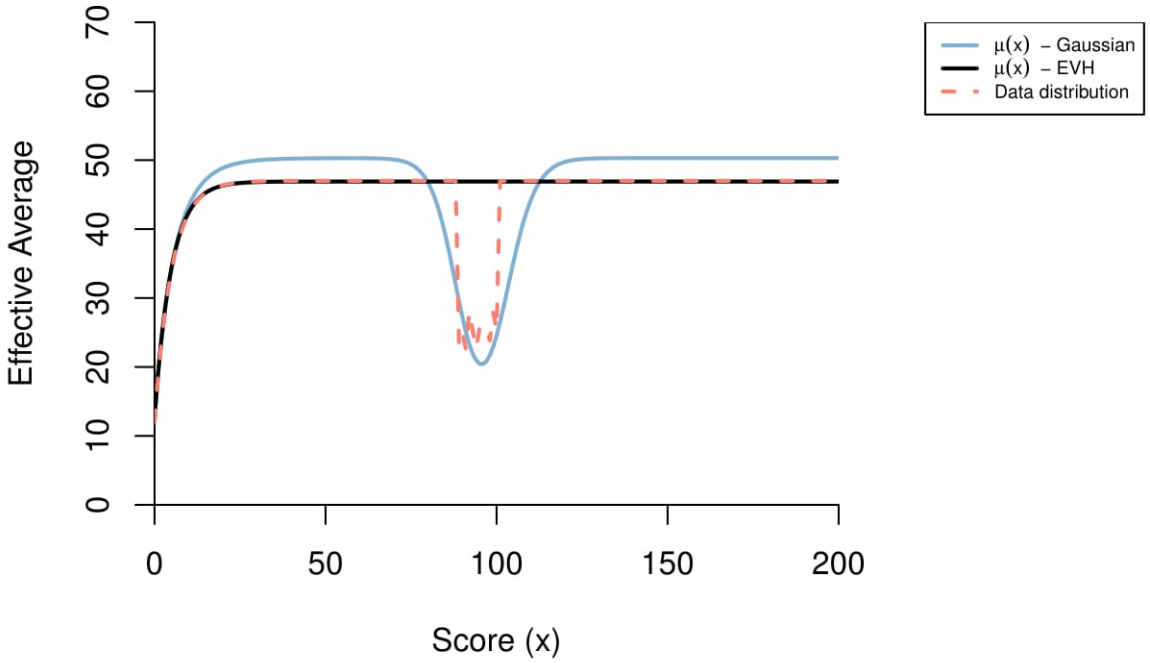


Figure 4.7. Predictive hazard function in terms of effective average, $\mu(x)$, for the fake data using the Gaussian hazard and exponential varying-hazard (EVH) models. The underlying data distribution that generated the fake data is overlaid.

in ability. The summaries suggest the Gaussian hazard model has compensated for the deviation in ability in the 90s by overestimating the value for the fake player's 'eye-in' ability, μ_2, and speed of transition between states, L, in comparison to the exponential varying-hazard model.

The AR(1) hazard model

The fake data set was then analysed using the AR(1) hazard model, producing another estimate for the fake player's effective average function. As anticipated, the estimate for the effective average curve under this model contains more fluctuations in ability than the other two models.

Figure 4.8 suggests that like the Gaussian hazard model, the AR(1) hazard model is also able to identify a significant decrease in ability during scores in the 90s. The model also appears to suggest that the fake player exhibits periods of increased batting ability at scores in the 70s and 80s and across a range of scores between 100 and 150.

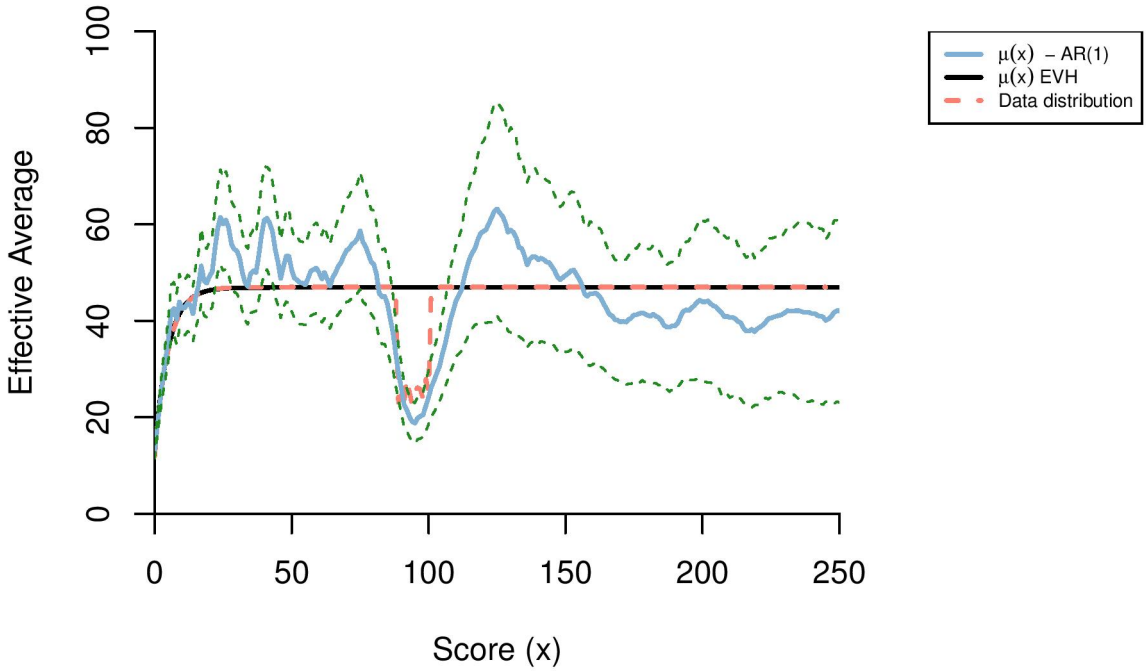


Figure 4.8. Predictive hazard function in terms of effective average, $\mu(x)$, for the fake data using the AR(1) hazard (including 68% credible intervals) and exponential varying-hazard models. The underlying data distribution that generated the fake data is overlaid.

Interestingly, the model also identifies several periods of increased ability during scores in the 20s and 40s, although these deviations are not as significant as the decrease in ability during the 90s.

To indicate whether or not these flexible models have successfully identified score-based temporal deviations in ability, the empirical hazard function for the fake data set was derived from the Kaplan-Meier estimator of the survival function (Kaplan & Meier, 1958).

A plot of the empirical hazard function suggests both the AR(1) and Gaussian hazard models have performed similarly in identifying score-based temporal deviations in batting ability. By definition, the empirical function becomes trivial at higher scores, as it eventually assigns a probability of dismissal of 1 at the player's highest score (assuming their high score is *not* a not out score), in this case, 245. However, it is clear from Figure 4.9 that both models are accurately identifying the temporal deviation in batting ability during the 90s.

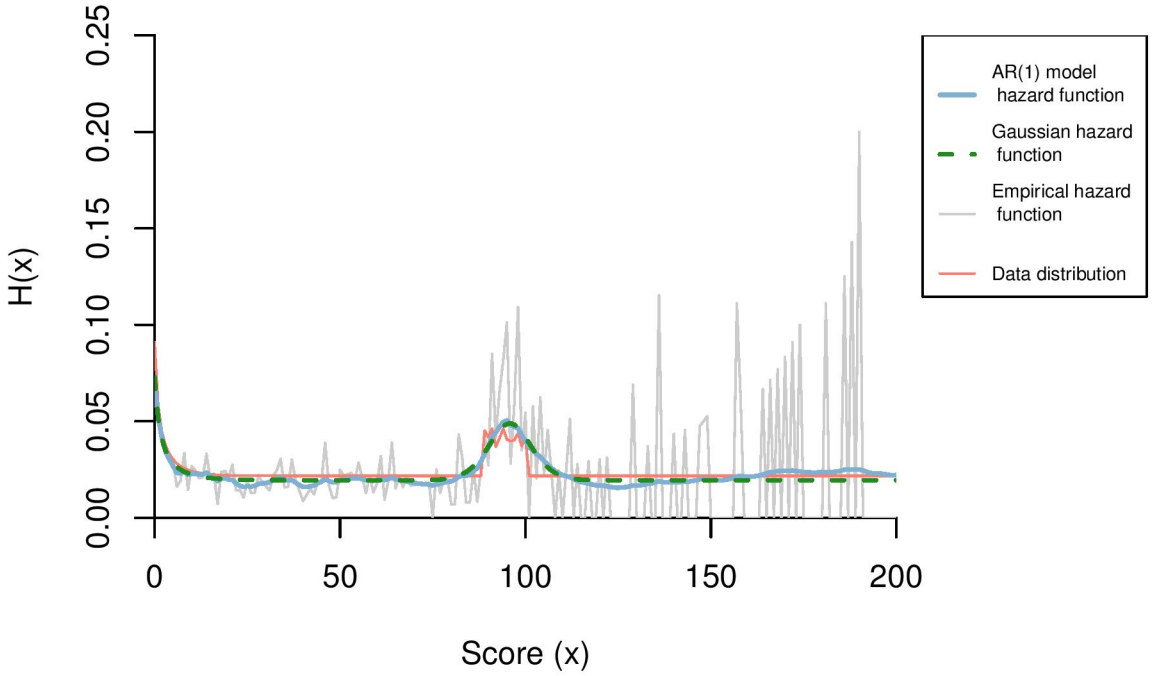


Figure 4.9. The empirical hazard function (grey) for the fake data with the estimated hazard function for the AR(1) and Gaussian hazard models overlaid.

The marginal likelihoods for each of the models fitted to the fake data are presented in Table 4.2. As expected the more flexible models are heavily favoured over the more rigid exponential varying-hazard model. With the highest marginal likelihood, the Gaussian hazard model appears to be the model most likely to apply to the data set, and assigns higher probability to it than the AR(1) hazard and exponential varying-hazard models. While the AR(1) hazard model can account for more flexibility in the data, it must assign some probability to outlandish and highly improbable datasets, resulting in a lower marginal likelihood.

Table 4.2: Marginal likelihoods for each of the three models fitted to the fake data

Model	$\log(Z)$
Gaussian hazard	-4114.06
AR(1) hazard	-4118.56
Exponential varying-hazard	-4129.01

4.4.2 Data

In order to reliably test each of the three models, a larger data set was required than both the small selection of retired players analysed in Chapter 2 and the New Zealand opening batsmen analysed in Chapter 3. As the models we have developed are focussed on batting, an appropriate data set of batsmen was selected for analysis under the assumptions of each of the three models.

The chosen data set consists of all international[3] players who have averaged more than 40 with the bat from at least 30 innings since the year 2000. In total, 47 players met this criteria, accounting for 7559 individual innings between them. The Test match batting records for each of the players in the data set are presented in Table B.1 in Appendix B.

In addition to these 47 players, former Australian opener Michael Slater was also analysed and presented as a case study, as a player who was notorious for being dismissed in the 90s. Slater played the majority of his career during the 1990s, and was dismissed in the 90s on 9 of the 23 occasions he scored at least 90 runs, making him an interesting prospect in the context of the Gaussian and AR(1) hazard models.

4.4.3 Analysis using the Gaussian hazard model

The Gaussian hazard model was run using the nested sampling algorithm for all players in the data set and samples were drawn from the posterior distribution, for each player. However, unlike the exponential varying-hazard model, it is not possible to simply present the posterior mean or median parameter estimates.

The main issue that arose when fitting effective average curves for the Gaussian hazard model, was selecting appropriate point estimates for k, ϕ and m. For the parameter m, most players tend to have one or two values that m gravitates towards, with the remainder of the posterior samples spreading m fairly evenly across the [0, 400] interval. Due to the wide posterior distribution for m, applying the approach of

[3] 'International' referring to countries with Test nation status (Australia, Bangladesh, England, New Zealand, Pakistan, South Africa, Sri Lanka, West Indies, Zimbabwe).

estimating the parameters by taking the posterior median, no longer provides valid results.

The posterior samples for m, for former English batsman Ian Bell are presented in Figure 4.10. It is clear that the Gaussian hazard model has inferred that Ian Bell likely exhibits some sort of temporal deviation in ability at scores in the 40s. However, if we were to take the posterior median of m for these samples, our estimate would suggest that Bell exhibits a deviation in ability at scores around 196, which is clearly not what the posterior distribution is indicating.

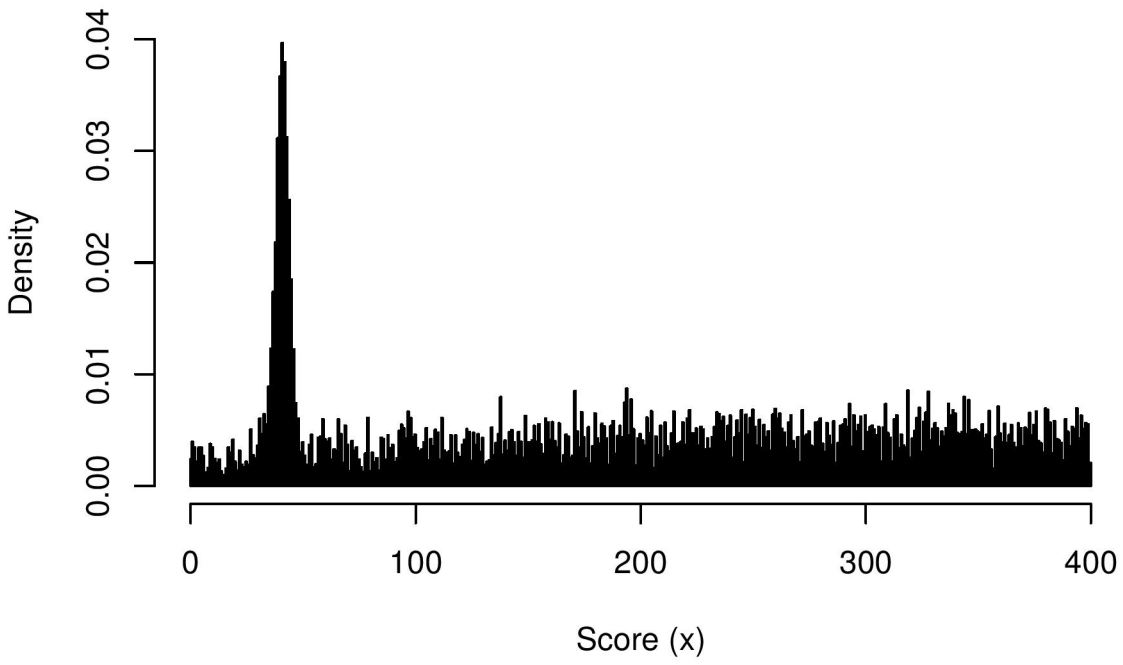


Figure 4.10. Histogram of posterior samples for m, for Ian Bell.

Furthermore, if we consider the posterior distribution for m, for former Sri-Lankan wicket-keeper and batsman, Kumar Sangakkara (Figure 4.11), we can see there are three ranges of scores where Sangakkara may exhibit a temporal deviation in ability ($\sim$70, $\sim$120, $\sim$170). However, from the marginal posterior distribution for m alone, it is impossible to know whether each of these score ranges are implying a decrease or increase in batting ability (i.e. positive or negative values for parameter k). Likewise, the deviation in ability across each score range will likely have unique widths (i.e.

varying estimates for the parameter ϕ).

The joint distribution for k and m, for Sangakkara (Figure 4.12), indicates that posterior samples with values $m \approx 70$, k tend to take positive values, suggesting a decrease in batting ability. On the contrary, for values $m \approx 120$ and $m \approx 170$, k tends to take negative values, indicating an increase in batting ability.

Similar disparities are present between ϕ and m, as seen in Figure 4.13. Posterior samples with values for $m \approx 70$, tend to have values for ϕ consistently clustered around 6. However, for values $m \approx 120$ and $m \approx 170$, ϕ appears to be spread across a much wider range of values. This makes it difficult to present any information learnt by the Gaussian hazard model in table form, as we can no longer simply take the posterior median for each model parameter. Instead, each player must be dealt with on a case-by-case basis.

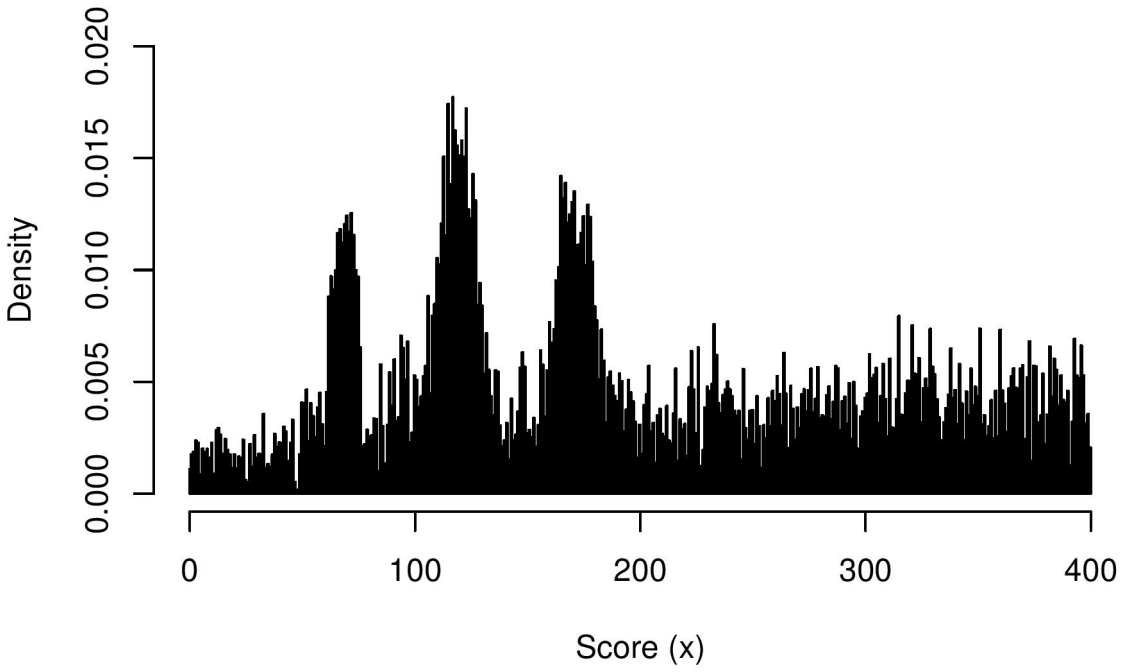


Figure 4.11. Histogram of posterior samples for m, for Kumar Sangakkara.

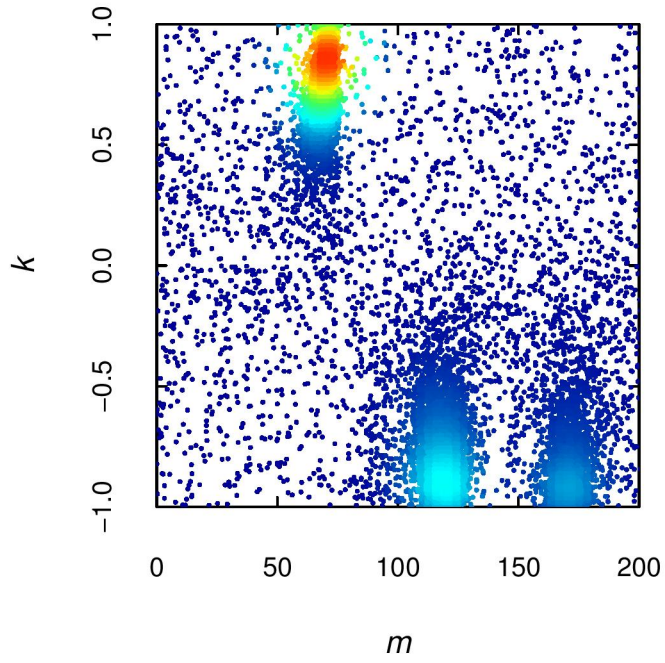


Figure 4.12. Joint posterior distribution for m and k, for Kumar Sangakkara.

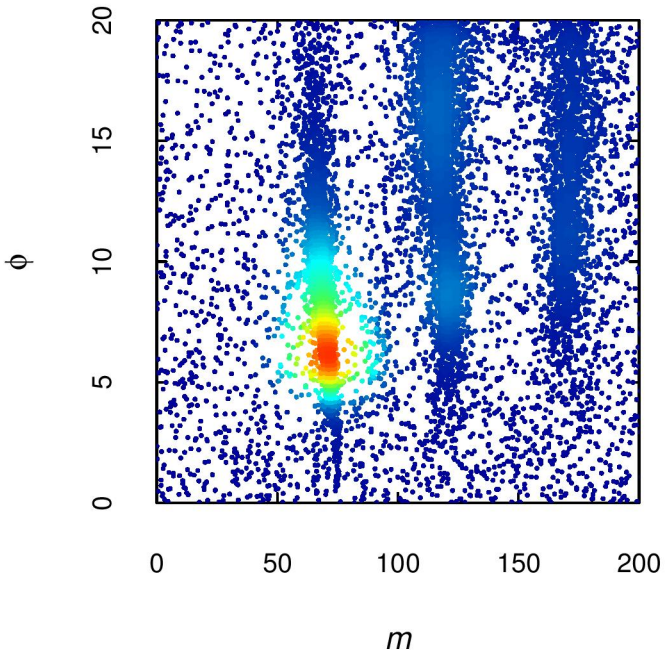


Figure 4.13. Joint posterior distribution for m and ϕ, for Kumar Sangakkara.

Predictive hazard functions

Rather than present the posterior inferences made by the Gaussian hazard model in a table, it is easier to see what the model has learnt from the data by plotting the predictive hazard function in terms of $\mu(x)$ for each player. These functions give us a much clearer idea of what the joint distributions for k, ϕ and m are inferring for each individual batsman.

Therefore, predictive hazard functions were fitted for the 15 players in the data set for whom the Gaussian hazard model was the most likely model to apply, based on the marginal likelihood. This group consists of five Australian and five English batsman, along with five other batsman from the rest of the world. The predictive hazard functions for each of these groups are presented in Figures 4.14, 4.15 and 4.16.

Additionally, each player was tested for whether or not they appear to exhibit a case of the 'nervous 90s' that adversely effects their batting ability. Equation 4.25 provides probability estimates for whether a given player appears to be batting worse when on scores in the 90s, in comparison to scores precluding and following the 90s. Estimates for each player's mean batting ability during the 50s, 60s, 70s, 80s and 100s were calculated for each posterior sample, and compared with their mean batting ability in the 90s. For example, to calculate the posterior probability that a player bats better in the 80s than in the 90s, we compute

$$P(\bar{\mu}_{80s} > \bar{\mu}_{90s}|\boldsymbol{d}) \approx \frac{1}{N}\sum_{i=1}^{N} I\{\bar{\mu}_{80s,i} > \bar{\mu}_{90s,i}\}, \tag{4.25}$$

for N posterior samples, where I is the indicator function that takes the value 1 if the mean batting ability in the 80s is higher than the mean batting ability in the 90s and 0 otherwise, for the i^{th} posterior sample. For comparison, prior probabilities were also computed and are presented alongside the posterior probabilities in Tables 4.3, 4.4 and 4.5.

It is also worth noting that the probability estimates in Equation 4.25 are calculated on the basis of a player batting *better* across a particular score range, compared

with in the 90s. For most players, there are a reasonable number of posterior samples that suggest equal ability between score ranges (this is sensitive to numerical precision). Therefore, the estimate $P(\bar{\mu}_{80s} > \bar{\mu}_{90s}|\boldsymbol{d}) = p$, certainly does not imply $P(\bar{\mu}_{80s} < \bar{\mu}_{90s}|\boldsymbol{d}) = 1 - p$.

Australian batsmen

The predictive hazard functions for the five Australian batsman for whom the Gaussian hazard model was the most likely model, are presented in Figure 4.14. Interestingly, it appears as though three players, Simon Katich, Damien Martyn and Mark Waugh, potentially experience a decline in batting ability as they approach a score of 100. Also of interest is the apparent transition speed between the initial and 'eye-in' batting abilities of Michael Clarke. Clarke does not appear to have reached an equilibrium batting ability until scoring at least 50 runs, suggesting he takes a very long time to completely get used to the conditions, but is a very good batsman once he has.

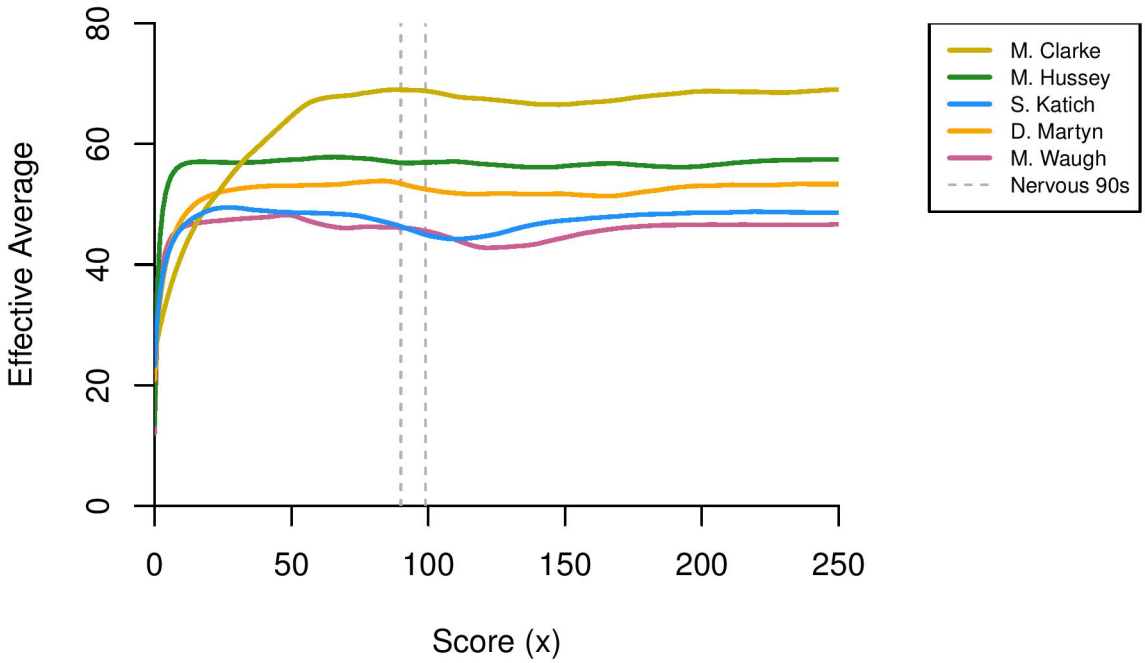


Figure 4.14. Predictive hazard functions for the Gaussian hazard model in terms of effective average, $\mu(x)$, for the Australian batsmen.

Table 4.3: Posterior probability estimates for the five Australian batsmen, comparing mean batting abilities during the 50s, 60s, 70s, 80s and 100s against the mean batting ability during the 90s.

Player	$\boldsymbol{P(\bar{\mu}_{50s} > \bar{\mu}_{90s}\|\boldsymbol{d})}$	$\boldsymbol{P(\bar{\mu}_{60s} > \bar{\mu}_{90s}\|\boldsymbol{d})}$	$\boldsymbol{P(\bar{\mu}_{70s} > \bar{\mu}_{90s}\|\boldsymbol{d})}$	$\boldsymbol{P(\bar{\mu}_{80s} > \bar{\mu}_{90s}\|\boldsymbol{d})}$	$\boldsymbol{P(\bar{\mu}_{100s} > \bar{\mu}_{90s}\|\boldsymbol{d})}$
M. Clarke	0.10	0.11	0.12	0.13	0.84
M. Hussey	0.21	0.22	0.22	0.22	0.30
S. Katich	0.36	0.36	0.35	0.35	0.49
D. Martyn	0.19	0.20	0.21	0.22	0.60
M. Waugh	0.38	0.39	0.39	0.38	0.23
Prior	0.17	0.17	0.17	0.17	0.56

The posterior probability estimates in Table 4.3 suggest there is little evidence to believe that Clarke's batting ability suffers from any nervousness while in the 90s.

Perhaps surprisingly, there is underwhelming support for the 'nervous 90s' significantly affecting batting ability, for both Simon Katich and Damien Martyn. From Figure 4.14, both players appear to be undergoing a decrease in batting ability as they approach 100. However, it is important to remember that the estimates for the mean batting abilities during the 50s, 60s, 70s, 80s, 90s and 100s are derived from the individual posterior samples, not the predictive hazard function. Given the insignificant posterior probability estimates for Katich and Martyn, we can conclude that most posterior samples estimate that both players bat better during the 90s than at earlier scores. However, a small number of posterior samples *do* exhibit significant decreases in batting ability around the 90s, dragging the predictive hazard function down for scores in the 90s.

Furthermore, there is no evidence to suggest that either Michael Hussey's or Mark Waugh's batting abilities suffer from the 'nervous 90s'. However, unlike Katich and Martyn, there does appear to be some weak evidence of a decrease in batting ability once passing 100 for each of these players, as indicated by $P(\bar{\mu}_{100s} > \bar{\mu}_{90s}|\boldsymbol{d})$. In terms of the predictive hazard functions, these estimates make reasonable sense, particularly in the context of Mark Waugh, whose predictive hazard function does appear to decrease immediately after scoring 100 runs.

English batsmen

Like the Australians, several English batsman appear to exhibit a decline in batting ability while on scores in the 90s, namely Andrew Strauss and Michael Vaughan. Oddly, both of these players also appear to peak in terms of batting ability just before entering the 90s. Conversely, Jonathan Trott appears to exhibit an increase in batting ability as he approaches a score of 100.

It comes as no surprise that Ian Bell appears to experience a brief period of increased ability at scores in the 40s, as the posterior distribution for parameter m in Figure 4.10 had a high density near these scores, perhaps indicating a real determination from Bell to reach the milestone of 50 runs once he gets within sight.

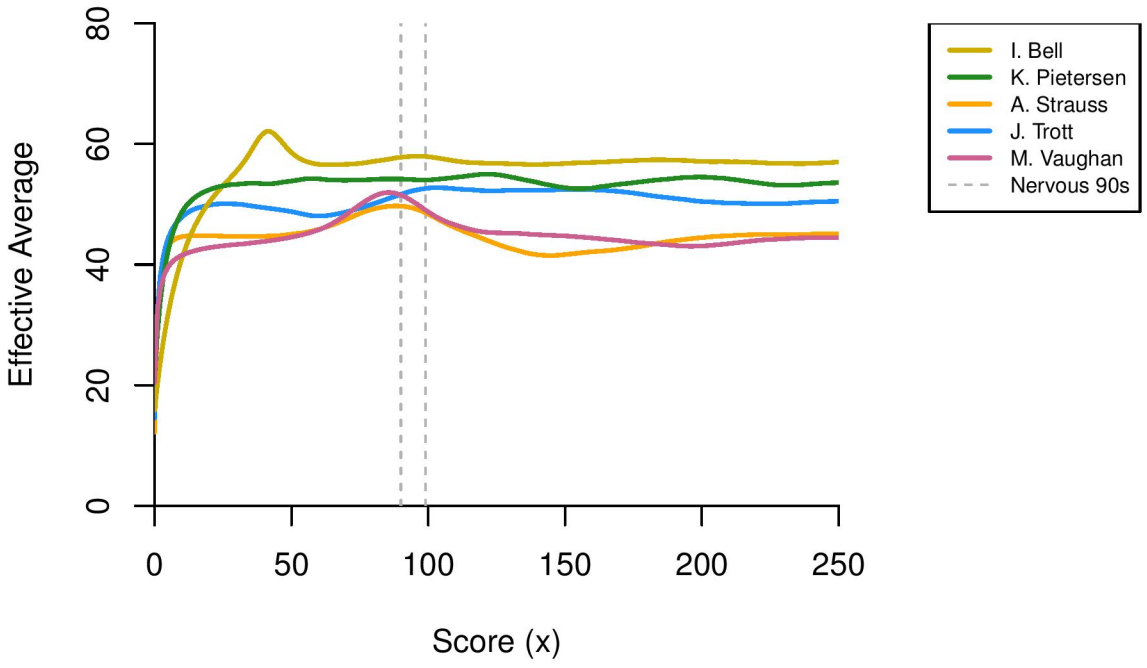


Figure 4.15. Predictive hazard functions for the Gaussian hazard model in terms of effective average, $\mu(x)$, for the English batsmen.

Again, like the Australians, there is no real evidence to suggest any of the English batsman are affected by the 'nervous 90s'. However, Andrew Strauss is similar to the likes of Mark Waugh, in the sense that there is some weak evidence to suggest he experiences a decrease in batting ability once passing 100 runs. This may be imply

Table 4.4: Posterior probability estimates for the five English batsmen, comparing mean batting abilities during the 50s, 60s, 70s, 80s and 100s against the mean batting ability during the 90s.

Player	$P(\bar{\mu}_{50s} > \bar{\mu}_{90s}\|d)$	$P(\bar{\mu}_{60s} > \bar{\mu}_{90s}\|d)$	$P(\bar{\mu}_{70s} > \bar{\mu}_{90s}\|d)$	$P(\bar{\mu}_{80s} > \bar{\mu}_{90s}\|d)$	$P(\bar{\mu}_{100s} > \bar{\mu}_{90s}\|d)$
I. Bell	0.19	0.16	0.13	0.11	0.88
K. Pietersen	0.11	0.12	0.13	0.13	0.74
A. Strauss	0.28	0.31	0.34	0.37	0.16
J. Trott	0.10	0.09	0.09	0.10	0.54
M. Vaughan	0.11	0.14	0.20	0.26	0.38
Prior	0.17	0.17	0.17	0.17	0.56

that Strauss and Waugh are more prone than others to losing their concentration, or playing a loose stroke, immediately after scoring a century.

Rest of world batsmen

The predictive hazard functions for the players from the rest of the world are presented in Figure 4.16. Here we can see the implications of the joint distribution for k and m, for Kumar Sangakkara (Figure 4.12), with three potential scoring areas causing deviations in batting ability. As expected, there appears to be a slight decrease in ability around scores of 70 for Sangakkara, followed by periods of increased ability near scores of 120 and 170.

Other players who appear to have meaningful score-based deviations in ability are South Africa's Ashwell Prince and Sri Lanka's Hashan Tillakaratne. Both players seem to bat better when on scores around 60-80, before returning to a state of relative equilibrium.

Figure 4.16 also contains the predictive hazard function for New Zealand opening batsman Mark Richardson, who was analysed in detail using the exponential varying-hazard model in Chapter 3. Richardson appears to undergo a gradual deterioration in ability, from the point of reaching a peak batting ability at a score of approximately 20, until reaching a score of 100.

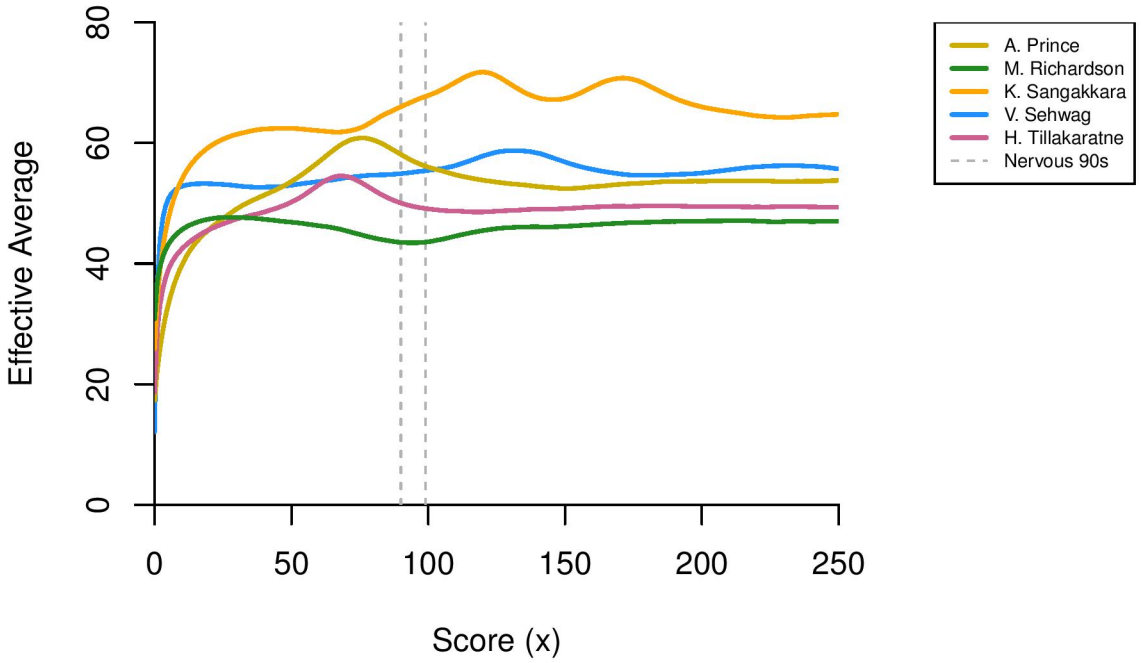


Figure 4.16. Predictive hazard functions for the Gaussian hazard model in terms of effective average, $\mu(x)$, for the rest of world batsmen.

Again, there is no significant evidence to suggest any players from the rest of world group experience any difficulty while batting on scores in the 90s. Unlike the Australian and English groups, there do not appear to be any players who exhibit a decrease in ability once passing 100.

Table 4.5: Posterior probability estimates for the five batsman from the rest the of world, comparing mean batting abilities during the 50s, 60s, 70s, 80s and 100s against the mean batting ability during the 90s.

Player	$\boldsymbol{P(\bar{\mu}_{50s} > \bar{\mu}_{90s}\|d)}$	$\boldsymbol{P(\bar{\mu}_{60s} > \bar{\mu}_{90s}\|d)}$	$\boldsymbol{P(\bar{\mu}_{70s} > \bar{\mu}_{90s}\|d)}$	$\boldsymbol{P(\bar{\mu}_{80s} > \bar{\mu}_{90s}\|d)}$	$\boldsymbol{P(\bar{\mu}_{100s} > \bar{\mu}_{90s}\|d)}$
A. Prince	0.16	0.21	0.25	0.26	0.67
M. Richardson	0.33	0.32	0.30	0.27	0.58
K. Sangakkara	0.03	0.03	0.03	0.04	0.85
V. Sehwag	0.06	0.07	0.08	0.09	0.42
H. Tillakaratne	0.24	0.27	0.28	0.28	0.50
Prior	0.17	0.17	0.17	0.17	0.56

4.4.4 Analysis using the AR(1) hazard model

Using diffusive nested sampling, posterior samples were generated for each of the 47 players in the data set. However, given the large number of parameters in the model, it is impractical to present the posterior parameter summaries for each player. Instead, to effectively convey the implications of the AR(1) hazard model, predictive hazard functions are presented for each of the batsmen for whom the AR(1) hazard model was the most likely of the three models, based on the marginal likelihood.

Predictive hazard functions

Of the 47 players in the data set, six were identified as having the AR(1) hazard model most likely fit their career Test match batting data. These six players have been split up and are presented in two groups of three, the first containing three Australian batsman, the second three batsman from the rest of the world.

Like with the Gaussian hazard model, each of these players was tested for whether or not their batting ability suffers while in the 'nervous 90s', using the results from Equation 4.25. The prior probabilities for these estimates are somewhat different under the assumptions of the AR(1) hazard model compared with the Gaussian hazard model. As the AR(1) hazard model allows for fluctuation in ability from score to score, there are practically zero instances of posterior samples estimating equal abilities across score ranges. Therefore, our prior probabilities indicate a player is more or less a half chance to be batting worse during the 90s compared with nearby score ranges.

Australian batsmen

Two opening batsman, Matthew Hayden and Chris Rogers, and middle order batsman Steve Waugh make up the three Australian batsmen for whom the AR(1) model was the most likely. Each of these players exhibit interesting score-based deviations in ability at various times during their innings.

Matthew Hayden appears to experience a period of difficulty during the 30s, how-

ever once he reaches 40, he begins to bat with increased ability. Interestingly, once Hayden hits a score of 90, his effective average appears to plummet, suggesting he may be a player whose batting ability suffers during the 'nervous 90s'.

The predictive hazard function for Chris Rogers suggests he experiences a significant decline in batting ability from scores just before 40, until just past the milestone of 50 runs. Within the scope of the AR(1) hazard model, Rogers maintains a fairly stable batting ability from a score of roughly 60 onwards. This may indicate Rogers tends to lose concentration after passing the milestone of 50, rather than 100 as observed for several players analysed under the Gaussian hazard model.

Compared with the other two Australians, Steve Waugh's predictive hazard function is fairly stable, other than a possible period of increased ability between scores of 110 and 150.

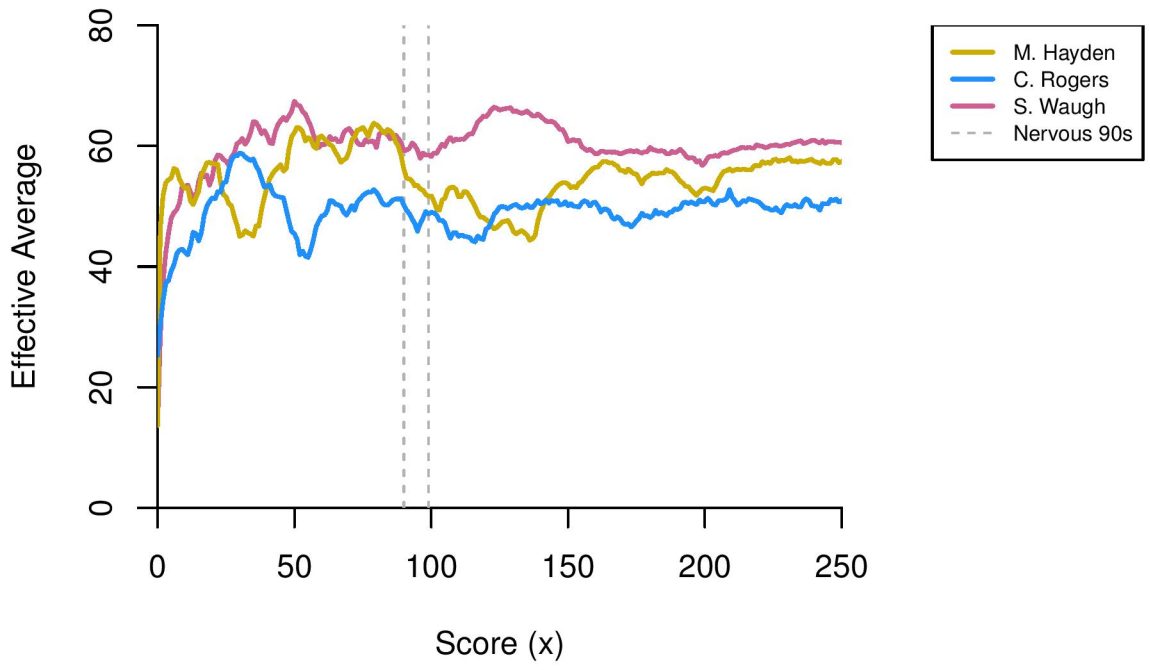


Figure 4.17. Predictive hazard functions for the AR(1) hazard model in terms of effective average, $\mu(x)$, for the Australian batsmen.

Table 4.6: Posterior probability estimates for the Australian batsmen, comparing mean batting abilities during the 50s, 60s, 70s, 80s and 100s against the mean batting ability during the 90s.

Player	$P(\bar{\mu}_{50s} > \bar{\mu}_{90s}\|\boldsymbol{d})$	$P(\bar{\mu}_{60s} > \bar{\mu}_{90s}\|\boldsymbol{d})$	$P(\bar{\mu}_{70s} > \bar{\mu}_{90s}\|\boldsymbol{d})$	$P(\bar{\mu}_{80s} > \bar{\mu}_{90s}\|\boldsymbol{d})$	$P(\bar{\mu}_{100s} > \bar{\mu}_{90s}\|\boldsymbol{d})$
M. Hayden	0.73	0.68	0.76	0.76	0.41
C. Rogers	0.34	0.50	0.57	0.57	0.47
S. Waugh	0.67	0.55	0.57	0.60	0.52
Prior	0.48	0.49	0.49	0.49	0.50

The posterior probability estimates from Table 4.6 tentatively confirm our suspicions regarding Matthew Hayden, that there is possible evidence (albeit very weak) that he does begin to bat worse once hitting the 90s. However, it is difficult to conclusively say that what Hayden experiences is exclusively due to the 'nervous 90s', as the decline in his effective average does not appear to stop once he passes 100. In fact, this deterioration in ability continues until Hayden reaches a score of roughly 140. Despite being a relatively quick scorer in terms of Test match cricket with a strike rate of 60.10, it is possible Hayden begins to suffer from either physical or mental fatigue once reaching a score as high as 90. Alternatively, as an opening batsman, Hayden may perceive his job of seeing off the opening bowlers and new ball as sufficiently achieved once he has scored 90 runs, and proceeds by taking the attack to the bowling side and playing more aggressively.

There is no evidence to suggest the batting abilities of either Chris Rogers or Steve Waugh are affected by the 'nervous 90s'. However, looking at scores leading up to, and following 50, suggests Rogers may be vulnerable to losing his wicket immediately after passing the milestone of 50 runs

$$P(\bar{\mu}_{40s,\ \text{Rogers}} < \bar{\mu}_{50s,\ \text{Rogers}}|\boldsymbol{d}) = 0.27.$$

Rest of world batsmen

Figure 4.18 depicts the predictive hazard functions for the three rest of world players, for whom the AR(1) hazard model was the most likely. Included in this group is

arguably the most famous modern-day cricketer, and undoubtedly one of the greatest batsmen of all-time, Indian great Sachin Tendulkar. During his heyday, Tendulkar had very few flaws in his game, however the AR(1) hazard model suggests if Tendulkar was especially vulnerable at any time during his innings (except for at the very beginning), it was between scores of 30 and 50. There is certainly no evidence to suggest Tendulkar's batting suffered from the 'nervous 90s', in fact, it appears as though his batting ability actually increases once he starts nearing 100. The other batsmen presented in Figure 4.18, Andy Flower and Matt Prior follow a similar trend of increasing in batting ability as they near 100.

Globally, there are cricketers and fans alike who believe Sachin Tendulkar was susceptible to the 'nervous 90s', as he holds the record for most dismissals in the 90s in Test cricket with 10 (an average of 1 every 32.9 innings). However, these beliefs are misguided, as the major reason Tendulkar holds this record is due to the fact he is by far the most capped player in history, having played in 200 Test matches (the

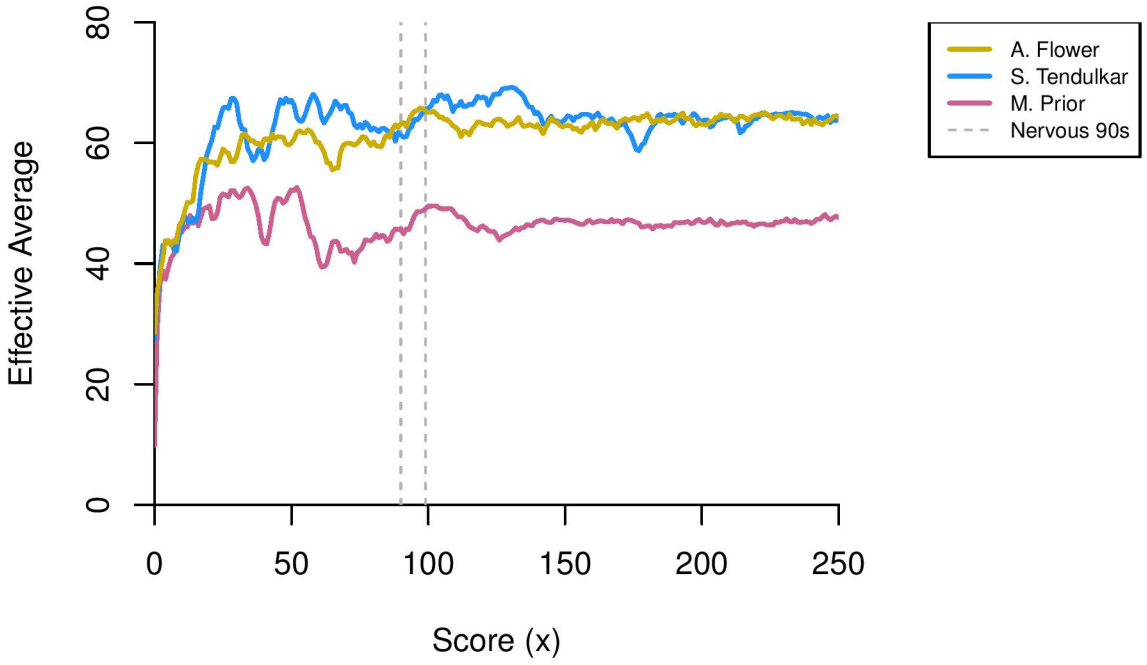


Figure 4.18. Predictive hazard functions for the AR(1) hazard model in terms of effective average, $\mu(x)$, for the rest of world batsmen.

Table 4.7: Posterior probability estimates for the batsman from the rest the of world, comparing mean batting abilities during the 50s, 60s, 70s, 80s and 100s against the mean batting ability during the 90s.

Player	$P(\bar{\mu}_{50s} > \bar{\mu}_{90s}\|\boldsymbol{d})$	$P(\bar{\mu}_{60s} > \bar{\mu}_{90s}\|\boldsymbol{d})$	$P(\bar{\mu}_{70s} > \bar{\mu}_{90s}\|\boldsymbol{d})$	$P(\bar{\mu}_{80s} > \bar{\mu}_{90s}\|\boldsymbol{d})$	$P(\bar{\mu}_{100s} > \bar{\mu}_{90s}\|\boldsymbol{d})$
A. Flower	0.35	0.23	0.30	0.35	0.50
M. Prior	0.55	0.29	0.32	0.37	0.54
S. Tendulkar	0.51	0.50	0.44	0.43	0.61
Prior	0.48	0.49	0.49	0.49	0.50

next highest are Australians Allan Border and Ricky Ponting with 168 caps). In fact, the player with the highest ratio of innings to dismissals in the 90s is Michael Slater (with 1 every 14.6 innings), who is analysed in Section 4.4.5.

The posterior probability estimates in Table 4.7 provide absolutely no evidence to support any cricketing folklore that Sachin Tendulkar is vulnerable to being dismissed in the 90s. Similarly, the probability estimates for both Andy Flower and Matt Prior provide no evidence to suggest either of these players were adversely affected by 'nervous 90s'.

4.4.5 Michael Slater: a case study

As a player with an abnormally high ratio of innings to times dismissed in the 90s, Michael Slater is often accused of suffering from the 'nervous 90s'. Slater is second only to Sachin Tendulkar in terms of number of dismissals in the 90s with 9, however achieved this number in far fewer innings. If we limit this statistic to only include innings where the player scored at least 90 runs, Tendulkar was dismissed just 10 times from 61 occasions (i.e. 16% of the time), while Slater was dismissed 9 times from 23 innings (39% of the time), a vastly inferior record. Therefore, the expectation was that if any player were to exhibit a clear-cut case of the 'nervous 90s', it would be Slater.

Posterior samples for Michael Slater's effective average were generated under the assumptions of each of the three models. The predictive hazard function for each

Table 4.8: Test career batting record for Michael Slater.

Player	**Matches**	**Innings**	**Not Outs**	**Runs**	**High-Score**	**Average**	**Strike Rate**	**100s**	**50s**
M. Slater	74	131	7	5312	219	42.83	53.29	14	21

Table 4.9: Marginal likelihoods for each of the three models fitted to the Michael Slater's Test career data.

Model	$\log(Z)$
Exponential varying-hazard	−590.13
Gaussian hazard	−590.05
AR(1) hazard	−589.85

model is plotted in terms of $\mu(x)$ in Figure 4.19. As indicated by Table 4.9, the more flexible Gaussian and AR(1) models appear to fit Slater's data better than the exponential varying-hazard model, suggesting Slater does exhibit some sort of temporal variation in ability over the course of an innings.

Interestingly, the predictive hazard functions for both the Gaussian hazard and AR(1) hazard models suggest Slater exhibits a period of increased batting ability during the 60s, 70s and 80s, rather than a significant decrease in ability during the 90s. The AR(1) hazard model also shows a slight decrease in batting ability during the 40s, right before the 50 run mark, potentially providing further evidence that Slater becomes unsettled before significant milestones.

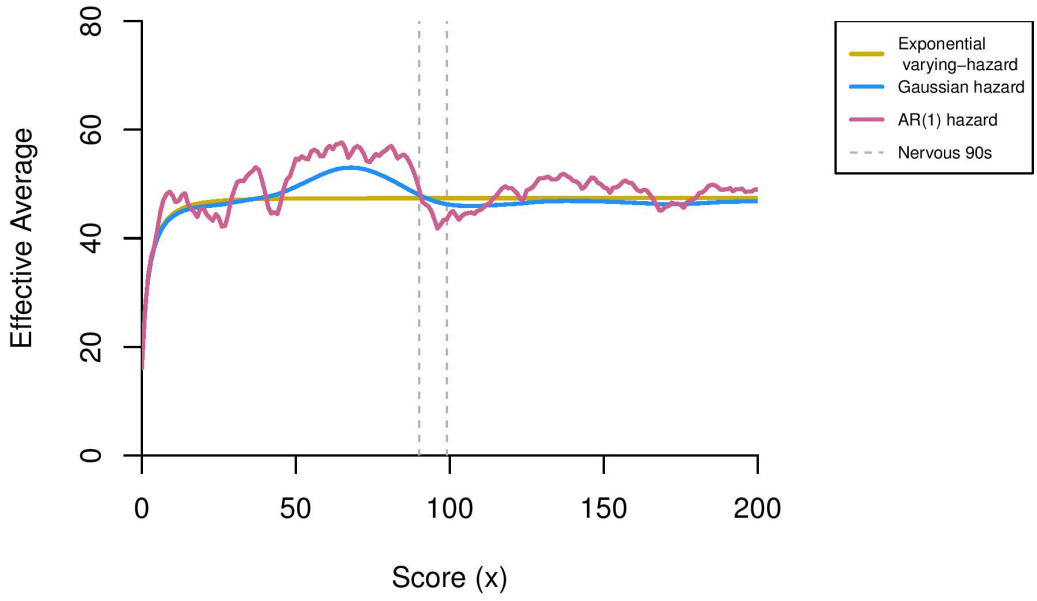


Figure 4.19. Predictive hazard functions for each of the three models in terms of effective average, $\mu(x)$, for Michael Slater.

Table 4.10: Posterior probability estimates for Michael Slater for each of the three models, comparing mean batting abilities during the 50s, 60s, 70s, 80s and 100s against the mean batting ability during the 90s. Prior probabilities are given for each model in red.

Model	$P(\bar{\mu}_{50s} > \bar{\mu}_{90s}\|\boldsymbol{d})$	$P(\bar{\mu}_{60s} > \bar{\mu}_{90s}\|\boldsymbol{d})$	$P(\bar{\mu}_{70s} > \bar{\mu}_{90s}\|\boldsymbol{d})$	$P(\bar{\mu}_{80s} > \bar{\mu}_{90s}\|\boldsymbol{d})$	$P(\bar{\mu}_{100s} > \bar{\mu}_{90s}\|\boldsymbol{d})$
Exponential varying-hazard	0.00	0.00	0.00	0.00	1.00
	(0.00)	(0.00)	(0.00)	(0.00)	(1.00)
Gaussian hazard	0.31	0.34	0.36	0.36	0.44
	(0.17)	(0.17)	(0.17)	(0.17)	(0.56)
AR(1) hazard	0.74	0.77	0.78	0.79	0.48
	(0.49)	(0.49)	(0.49)	(0.49)	(0.50)

The prior and posterior probabilities for the exponential varying-hazard model are either 0 or 1 due to the fact that the effective average is a monotonically increasing function. Unsurprisingly, this implies that this model is unable to pick up on any temporal deviation in ability, outside of the initial period of a player getting their 'eye-in'.

As the most likely model to apply to Slater's data, we are primarily interested in the findings of the AR(1) hazard model. The posterior probability estimates do suggest there is some evidence that Slater begins to bat worse, once he enters the 90s. However, the very notion a player suffering from the 'nervous 90s', implies that a batsman will begin to bat better once they pass the 100 run mark, which is not supported in Slater's case. This leads us to conclude that rather than being adversely affected by the 'nervous 90s', it is possible that Slater simply bats very well for a period of time after passing 50 and consequently reaches scores of 90 more frequently than a player of his calibre should.

Prior and posterior probabilities comparing Slater's batting abilities in the 30s and 50s with ability during the 40s are shown in Table 4.11. While there is some evidence to suggest Slater bats worse during the 40s, these estimates are hardly significant shifts away from the prior probabilities.

Therefore, despite being a player whose data suggests a possible decline in batting ability when nearing the mark of 100 runs, there is only weak evidence to suggest that Michael Slater begins to bat worse once he enters the 90s. It is unlikely that

Table 4.11: Posterior probability estimates for Michael Slater for the AR(1) hazard model, comparing mean batting abilities during the 30s and 50s against the mean batting ability during the 40s.

Model	$\boldsymbol{P(\bar{\mu}_{30s} > \bar{\mu}_{40s}\|d)}$	$\boldsymbol{P(\bar{\mu}_{50s} > \bar{\mu}_{40s}\|d)}$
AR(1) hazard	0.60	0.73
Prior	0.46	0.53

this reduction in batting ability during the 90s can be attributed to nerves, as Slater does not appear to bat much better even once passing the 100 run mark. Instead, the Gaussian hazard and AR(1) hazard models both suggest it is more realistic that Slater bats particularly well immediately after passing the milestone of 50 runs.

4.5 Limitations and conclusions

As seen from the predictive hazard functions in Section 4.4.3, the Gaussian hazard model allows for some score-based variation in batting ability. While the exponential varying-hazard model's strength was identifying how well a player bats when they first arrive at the crease, and how much better they become, the Gaussian hazard model focusses on identifying score ranges that exhibit increased or decreased ability. These score ranges typically occur at scores after the batsman is deemed to have their 'eye-in'.

Looking specifically at player batting ability across the range of scores in the 90s, the Gaussian hazard model does not conclusively identify any players whose batting ability is affected by the 'nervous 90s'. However, the model has identified several players who are possibly guilty of losing their concentration after passing the significant milestone of 100 runs (perhaps evidence of the *'hazardous 100s'*?).

Other than the limitations identified by the exponential varying-hazard model (see Section 2.4), a drawback of the Gaussian hazard model is that only one period of temporal deviation in ability is allowed under the definition of the model. For batsmen such as Kumar Sangakkara, the predictive hazard function suggests there are multiple possible periods of deviation in batting ability, however, only a single

period can be fitted under the Gaussian hazard model. This is the sort of problem we hoped to solve by fitting the AR(1) hazard model in section 4.4.4.

While the AR(1) hazard model allows for far more score-based variation in batting ability than both the exponential varying-hazard and Gaussian hazard models, it was the most likely model for only six of the 47 batsmen in the dataset. Our findings suggest that of these six players, only Matthew Hayden appeared to exhibit a possible decline in batting ability during the 'nervous 90s', although the posterior probabilities are not overly convincing. These posterior probabilities also indicated that Chris Rogers is potentially vulnerable immediately after passing the milestone of 50 runs (a possible example of the '*fallible 50s*'?), but again the evidence is not overwhelmingly strong.

Therefore, given the relative preference for the Gaussian hazard model over the AR(1) hazard model for players who *do* exhibit temporal variation in ability, an adaptation to the Gaussian model could be made. Including the ability to fit multiple Gaussian functions within the Gaussian hazard model, would allow for multiple periods of temporal variation in batting ability to be fitted, for each individual batsman. Multiple Gaussian functions may better describe a player's batting ability over the course of an innings, better than an autoregressive process.

Considering the special case of Michael Slater, who was dismissed in the 90s on an unusually large number of occasions, both the Gaussian hazard and AR(1) hazard models identified an increase in batting ability for scores leading up to 90, rather than a significant decrease in ability during the 90s. This leads us to conclude that Slater was likely a better batsman than expected while on scores after 50, rather than a bundle of nerves while in the 90s.

The Gaussian hazard and AR(1) hazard models were the most likely to apply for 21 of the 48 players in the data set, with the exponential varying-hazard model the most likely to apply for the remaining 27 players. Predictive hazard functions for these players are presented in Appendix B. To thoroughly address the overall fit of each of the three models, a more comprehensive exercise in model comparison is carried out in Chapter 5.

Chapter 5

Marginal likelihoods and model comparison

5.1 Overview

While the models presented in Chapter 4 allow for more complex interactions between a batsman's score and batting ability than the initial exponential varying-hazard model from Chapter 2, the question still remains: which model is the 'best'?

In this chapter, the fits of the three models are compared for each of the international batsmen analysed in Chapter 4. Although there is plenty of uncertainty pertaining to the temporal variations in ability exhibited by many of the players, it is possible to make inferences as to how like each model is to apply to the data of world-class batsmen.

A posterior distribution for determining the probability of a model given the data is derived in Section 5.2. Using this distribution, we can estimate what proportion of world-class batsmen undergo periods of score-based deviation in ability throughout their innings.

Finally, in Section 5.3, the thesis is summarised by calculating the overall marginal likelihood of the data, given the assumptions of the three models that have been fitted to the data set of international batsmen.

5.2 Measuring the undetectable

As mentioned throughout this thesis, using nested sampling to fit each model allows us to easily calculate the marginal likelihood or evidence, Z. This allows for model selection to be carried out trivially when comparing models for an individual player, by simply using Bayes factor (Equation 1.4). For example, comparing the exponential varying-hazard and AR(1) hazard models for Michael Slater gives

$$\frac{\exp(-589.85)}{\exp(-590.13)} = 1.32,$$

indicating the AR(1) hazard model is favoured by a factor of approximately 1.3 to 1. However, it is not quite so straightforward comparing models across all players in the data set.

Table B.2 in Appendix B presents the marginal likelihoods when fitting each of the three models, to each player in our data set of interest. Of the 47 players, the exponential varying-hazard model was the most likely for 26, the Gaussian hazard model for 15 and the AR(1) hazard model for 6. Interestingly, this implies that for the majority of players, the least flexible model was the most likely model to apply to the data. This observation can best be explained by the principle of Occam's razor, that one should accept the simplest explanation that fits the data (Thorburn, 1918; Jefferys & Berger, 1992; MacKay, 2003). That is, if a player does not appear to exhibit a significant amount of temporal variation in batting ability during their innings, then we have no reason to believe they do. Including additional parameters that allow for deviations in batting ability only adds needless complication. Therefore, the best model to fit to such a player, is one that does not allow for deviations in ability once a player has their 'eye-in', i.e. the exponential varying-hazard model.

Looking at the marginal likelihoods in Table B.2 from another perspective tells us the exponential varying-hazard model was the *worst* fit for 7 players, the Gaussian hazard model for 3 and the AR(1) hazard model for 37. Regardless of perspective, the findings consistently suggest the AR(1) hazard model tends to explain a player's

data the least well. However, from this angle it is more difficult to clearly determine which model is the 'best', between the exponential varying-hazard model and the Gaussian hazard model.

While we cannot be certain about the model fits for each individual player, we can use the marginal likelihoods to make inferences regarding the wider population of modern-day world-class batsmen. Using a technique referred to as 'measuring the undetectable' (Lang et al., 2009), we can derive a posterior distribution, which summarises the probability of each model being the most likely model to apply to the data of a world-class batsman.

Deriving the posterior distribution for model comparison

Let $\omega_i \in \{1, 2, 3\}$ be the proposition that each model is the most likely to apply to the data for the i^{th} player, where $\omega = 1$ implies the exponential varying-hazard model is most likely, $\omega = 2$ implies the Gaussian hazard model, and $\omega = 3$ implies the AR(1) hazard model. Also, define q_j as the conditional probability that model j is the most likely model to apply to player i, where $j = \{1, 2, 3\}$, giving the vector $\boldsymbol{q} = \{q_1, q_2, q_3\}$. Therefore, we can write the conditional distribution for ω_i as

$$p(\omega_i|\boldsymbol{q}) = \{q_1, q_2, q_3\} \tag{5.1}$$

where all observations ω_i, are assumed independent and identically distributed. It is worth noting we have already computed Z_{ij}, the marginal likelihoods for each model, for each player.

If $\boldsymbol{d_i}$ is the data for the i^{th} player in our data set, and $\boldsymbol{d}$ is the vector of data sets for all 47 players, our goal is to derive a posterior distribution for $p(\boldsymbol{q}|\boldsymbol{d})$, which will infer the probability that each model is the most likely for a given player, as the hyperparameter $\boldsymbol{q}$ will determine the probability that model 1, 2 or 3 applies to a particular player. Instinctively, to derive $p(\boldsymbol{q}|\boldsymbol{d})$, we might turn to Bayes' theorem

$$p(\boldsymbol{q}|\boldsymbol{d}) = \frac{p(\boldsymbol{q})\ p(\boldsymbol{d}|\boldsymbol{q})}{p(\boldsymbol{d})}$$

$$\propto p(\boldsymbol{q})\ p(\boldsymbol{d}|\boldsymbol{q}) \tag{5.2}$$

Working backwards and applying the product rule we can rewrite this as

$$\begin{aligned} p(\boldsymbol{q}|\boldsymbol{d}) &\propto p(\boldsymbol{q}) \int p(\boldsymbol{d}, \boldsymbol{\omega}|\boldsymbol{q})\ d\boldsymbol{\omega} \\ &= p(\boldsymbol{q}) \int p(\boldsymbol{\omega}|\boldsymbol{q})\ p(\boldsymbol{d}|\boldsymbol{\omega}, \boldsymbol{q})\ d\boldsymbol{\omega} \\ &= p(\boldsymbol{q}) \int \left[\prod_{i=1}^{47} p(\omega_i|\boldsymbol{q})\right] p(\boldsymbol{d}|\boldsymbol{\omega})\ d\boldsymbol{\omega} \\ &= p(\boldsymbol{q}) \int \prod_{i=1}^{47} p(\omega_i|\boldsymbol{q})\ p(\boldsymbol{d_i}|\omega_i)\ d\omega_1, ..., d\omega_{47} \end{aligned} \tag{5.3}$$

Given our assumption that ω_i and $\boldsymbol{d_i}$ are conditionally independent (which is not perfect, but is the best available option for now), we can pull out the product term and treat the integrand as a summation over all models, over all ω_i, giving

$$p(\boldsymbol{q}|\boldsymbol{d}) = p(\boldsymbol{q}) \prod_{i=1}^{47} \left[\sum_{j=1}^{3} p(\omega_i|q_j)\ p(\boldsymbol{d_i}|\omega_i = j)\right], \tag{5.4}$$

where the conditional distribution for $p(\omega_i|q_j) = q_j$ is given by Equation 5.1, and $p(\boldsymbol{d_i}|\omega_i = j)$ is simply the marginal likelihood Z_{ij} for each model.

Since the nested sampling algorithm was defined to calculate the log-likelihood (or log-evidence), using MCMC we evaluate the log-posterior

$$\log\left[p(\boldsymbol{q}|\boldsymbol{d})\right] = \log\left[p(\boldsymbol{q})\right] + \sum_{i=1}^{47}\sum_{j=1}^{3} \log\left[p(\omega_i|q_j)\ p(\boldsymbol{d_i}|\omega = j)\right]. \tag{5.5}$$

Graphically, the model structure can be presented as a directed acyclic graph, as in Figure 5.1.

To compute the posterior marginal distribution for $\boldsymbol{q} = \{q_1, q_2, q_3\}$, we must first specify a prior distribution for $\boldsymbol{q}$. As it is a vector of probabilities, $\boldsymbol{q}$ was assigned a Dirichlet prior with concentration parameters equal to 1, giving the expected value of $\boldsymbol{q} = \{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$. For computational efficiency, values for $\boldsymbol{q}$ were sampled in the MCMC

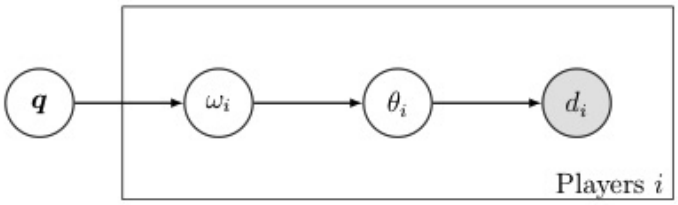


Figure 5.1. Directed acyclic graph illustrating the structure of the model.

algorithm by sampling values $\boldsymbol{e} = \{e_1, e_2, e_3\}$ from an Exponential(1) distribution; $\boldsymbol{q}$ is obtained by normalising $\boldsymbol{e}$ (i.e. $\boldsymbol{q} = \frac{\boldsymbol{e_j}}{\sum_j \boldsymbol{e}}$).

The posterior marginal distribution for $\boldsymbol{q}$ is shown in Figure 5.2, indicating that the exponential varying-hazard model is the model most likely to apply to a world-class batsman's data, followed by the Gaussian hazard model, with the AR(1) hazard model being the least likely. Therefore, our updated state of knowledge regarding $\boldsymbol{q}$ can be obtained by taking the posterior means and is summarised as $\hat{\boldsymbol{q}} = \{0.49, 0.40, 0.11\}$. These are also our posterior predictive probabilities for the 'next' world-class batsman. As $q_2 + q_3 > 0.5$, we would predict that there is just over a half chance that the next world-class batsman analysed by the three models, would exhibit some form of significant temporal variation in batting ability, during their innings.

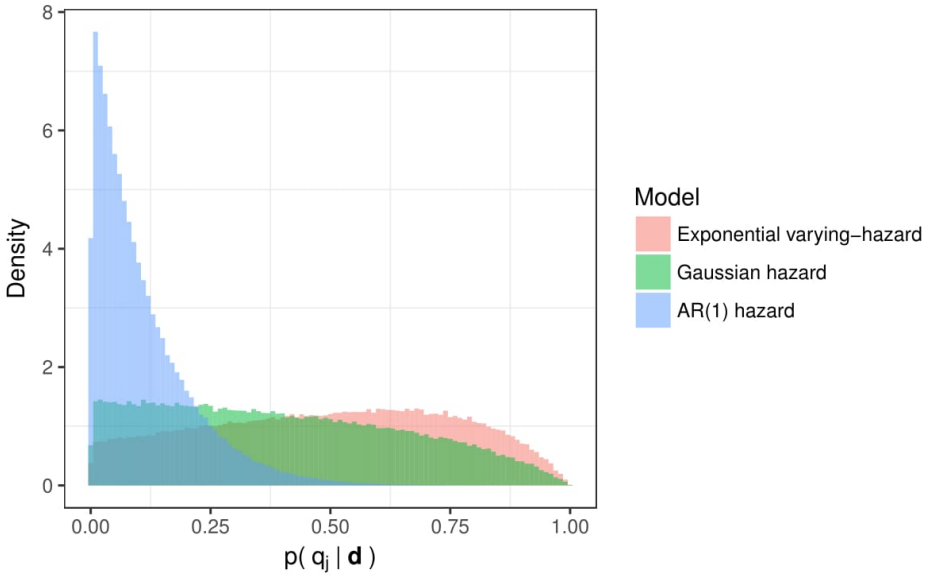


Figure 5.2. Histograms representing the posterior marginal distributions for q_j, indicating the probability of each model being the most likely to apply to a world-class batsman's career data.

5.3 Marginal likelihood of the thesis

As we have computed the likelihoods for each model, for every player, the entirety of this thesis can be summarised using just a single marginal likelihood value, representing the likelihood of the data, given the assumptions of the three models that have been fitted. Consider the proposition, T, that for each player in the data set, one of the three models applies. Then, the most computationally convenient form of calculating the marginal likelihood of observing the data, $\boldsymbol{d}$, given T is

$$\begin{aligned} p(\boldsymbol{d}|T) &= \int p(\boldsymbol{q})\ p(\boldsymbol{d}|\boldsymbol{q})\ d\boldsymbol{q} \\ &\approx \text{Average value of } p(\boldsymbol{d}|\boldsymbol{q}) \text{ when sampling from } p(\boldsymbol{q}). \end{aligned} \tag{5.6}$$

Using the result of Equation 5.6 we can use MCMC to compute the marginal likelihood for the data set of 47 international players. The corresponding log-likelihood value that summarises this thesis is -34121.21. This value alone has little meaning but allows for different model assumptions to be compared through the ratio of marginal likelihood values, or Bayes' factors (Skilling, 2006). Therefore, assuming the exact same data set of 47 international batsmen is used, it is easy to compare the performance of another set of models with the exponential varying-hazard, Gaussian hazard and AR(1) hazard models.

Chapter 6

Concluding statements and further work

This thesis has presented new, innovative methods of quantifying how well cricket players bat over the course of an innings. As expected, the models developed provide conclusive evidence that batsmen do not bat with equal ability throughout their innings. Rather, it takes time to get used to the specific match conditions, supporting the cricketing notion of 'getting your eye-in'.

The exponential varying-hazard model detailed in Chapter 2 enables the identification of (1) how well a batsman performs when they first begin an innings, (2) how much better they perform once they have their 'eye-in' and (3) how long it takes them to transition between their initial and 'eye-in' batting abilities. Despite ignoring important variables, such as balls faced or minutes batted, which may be included in future models, these tools can provide coaches and players with invaluable information as to which players in both their own, and opposition teams, are particularly vulnerable at the beginning of their innings. This may allow for the identification of particular players who are more or less suited to open the batting, leading to practical implications in terms of batting order and team selection policy. Additionally, captains can use this information to set more attacking fields for longer, for opposition players who exhibit a prolonged period of vulnerability at the start of their innings.

Applying the model in a hierarchical structure allows for inference to be made regarding a wider group of players, as well as being able to make informed predictions concerning the abilities of the next player to join the cohort of players analysed. In the case of New Zealand opening batsmen, the hierarchical model confirms our suspicions that opening has been a position of concern for the national side since the year 2000. The relatively low estimates for initial batting abilities, μ_1, suggest our concept of opening batsmen being more 'robust' than other batsman is not widely supported, although this may be due to the talent pool focussed on. Applying the hierarchical model to a country that has produced a larger number of world-class opening batsmen in the recent past (e.g. Australia, England, India, South Africa), may yield a different conclusion.

Developing the exponential varying-hazard model further to allow for temporal deviations in ability, other than at the beginning of a player's innings, allows for the exploration of popular cricketing superstitions, such as the 'nervous 90s'. However, the findings from the Gaussian hazard and AR(1) hazard model in Chapter 4 suggest very few players exhibit a significant decline in ability during the 90s, and those players who do, do not appear to immediately improve in terms of batting ability once scoring 100. Realistically, no players have played Test cricket for long enough to confidently confirm or refute the existence of any detrimental effects due to the 'nervous 90s'. This implies that rather than assuming a player's batting ability deteriorates during the 'nervous 90s', players should be presumed to be unaffected by the 'nervous 90s', unless proven otherwise, akin to the legal right to be presumed innocent until proven guilty. In fact, the most common score-based deviations in ability appear to occur immediately *after* passing significant milestones, giving rise to the sentiments of the *'fallible 50s'* and the *'hazardous 100s'*.

While it is difficult to confirm the existence of any systematic score-based deviations in ability between players, the Gaussian and AR(1) hazard models still provide evidence of temporal variation in batting ability pertaining to individual players. In terms of practicality, this may supply the fielding team captain with knowledge as to when the best times to attack and defend are, during an opposing player's in-

nings. A traditional fielding approach would see a captain set attacking fields during the early stages of a player's innings, with fields becoming more and more defensive as the batsman scores more runs. However, the results of the Gaussian and AR(1) hazard models may offer the fielding side additional attacking opportunities, during periods where a batsman exhibits a period of decreased ability that would usually go unnoticed.

As both the Gaussian and AR(1) hazard models suggest the presence of temporal variation in batting ability *during* an innings, future models could explore the plausibility of variation existing *between* innings. Such fluctuations in ability may exist on two scales: long-term variation due to factors such as age and experience, and short-term variation due to opposition, local pitch and weather conditions, player form and player fitness.

Maintaining records concerning the performance of batsmen in certain pitch and weather conditions, may help with team selection in foreign batting conditions (e.g. teams such as Australia and New Zealand travelling to the sub-continent). While most international sides likely maintain some data of this nature, the continual failure of certain players in particular conditions suggests that teams could be doing far more in the way of pitch and weather analysis. This is especially true for countries with deep talent pools, which should allow for more flexibility in selecting their optimal starting XIs.

Regardless of local conditions, it is not uncommon to see the performances of older, respected players, decline as they approach the end of their career. Selectors are often faced with the difficult decision of dropping the older player in favour of somebody younger — and possibly angering the public and fanbase — or sticking with the older player who is producing sub-optimal results. Similarly difficult decisions arise when, a regular player returns from injury, but the player who has temporarily replaced them is performing extremely well. Allowing for more complex relationships between the parameters of interest will give our models the ability to answer these more difficult questions.

The aim of this thesis has not been to reinvent the coaching manual to focus en-

tirely on the use of statistical models in cricket. However, developing models which can accurately account for variation in ability due to long-term factors such as age, or short-term factors such as pitch conditions and fitness, provide coaching and management staff with the information to make more informed decisions. Ultimately, professional cricket is a results-driven industry. The development of any statistical techniques that effectively model player ability, will provide teams with more strategic tools at their disposal, which in close matches, can be the difference between winning and losing.

References

Agresti, A., & Kateri, M. (2011). *Categorical data analysis*. Springer.

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology*, *13*(1), 61–98.

Bailey, M., & Clarke, S. R. (2006). Predicting the match outcome in one day international cricket matches, while the game is in progress. *Journal of Sports Science and Medicine*, *5*(4), 480.

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2014). Julia: A fresh approach to numerical computing. *arXiv preprint arXiv:1411.1607*.

Bracewell, P. J., & Ruggiero, K. (2009). A parametric control chart for monitoring individual batting performances in cricket. *Journal of Quantitative Analysis in Sports*, *5*(3).

Brewer, B. J. (2008). Getting your eye in: A Bayesian analysis of early dismissals in cricket. *ArXiv preprint:0801.4408v2*.

Brewer, B. J. (2014). *Bayesian inference and computation: a beginners guide.* Retrieved from `https://www.stat.auckland.ac.nz/~brewer/wsbook.pdf`

Brewer, B. J., & Elliott, T. M. (2014). Hierarchical reverberation mapping. *Monthly Notices of the Royal Astronomical Society: Letters*, *439*(1), L31–L35.

Brewer, B. J., & Foreman-Mackey, D. (2016). Dnest4: Diffusive nested sampling in c++ and python. *arXiv preprint arXiv:1606.03757*.

Brewer, B. J., Pártay, L. B., & Csányi, G. (2011). Diffusive nested sampling. *Statistics and Computing*, *21*(4), 649–656.

Brooker, S., & Hogan, S. (2011). *A method for inferring batting conditions in ODI cricket matches from historical data.* Working Papers in Economics 11/44, University of Canterbury, Department of Economics and Finance. Retrieved from `http://www.econ.canterbury.ac.nz/RePEc/cbt/econwp/1144.pdf`

Brooks, R. D., Faff, R. W., & Sokulsky, D. (2002). An ordered response model of test cricket performance. *Applied Economics*, *34*(18), 2353–2365.

Cai, T., Hyndman, R. J., & Wand, M. (2002). Mixed model-based hazard estimation. *Journal of Computational and Graphical Statistics*, *11*(4), 784–798.

Carter, M., & Guthrie, G. (2004). Cricket interruptus: fairness and incentive in limited overs cricket matches. *Journal of the Operational Research Society*, *55*(8), 822–829.

Caticha, A., & Giffin, A. (2006). Updating probabilities. In *AIP conference proceedings* (Vol. 872, pp. 31–42).

Clarke, S. R. (1988). Dynamic programming in one-day cricket-optimal scoring rates. *Journal of the Operational Research Society*, *39*(4), 331–337.

Clarke, S. R., & Norman, J. M. (1999). To run or not?: Some dynamic programming models in cricket. *Journal of the Operational Research Society*, *50*(5), 536–545.

Clarke, S. R., & Norman, J. M. (2003). Dynamic programming in cricket: Choosing a night watchman. *Journal of the Operational Research Society*, *54*(8), 838–845.

Damodaran, U. (2006). Stochastic dominance and analysis of ODI batting performance: the Indian cricket team, 1989-2005. *Journal of Sports Science and Medicine*, *5*(4), 503–508.

Davis, J., Perera, H., & Swartz, T. B. (2015). A simulator for Twenty20 cricket. *Australian and New Zealand Journal of Statistics*, *57*(1), 55–71.

Duckworth, F. C., & Lewis, A. J. (1998). A fair method for resetting the target in interrupted one-day cricket matches. *Journal of the Operational Research Society*, *49*(3), 220–227.

Elderton, W., & Wood, G. H. (1945). Cricket scores and geometrical progression. *Journal of the Royal Statistical Society*, *108*(1/2), 12–40.

Foreman-Mackey, D. (2016, jun). corner.py: Scatterplot matrices in Python. *JOSS*, *1*(2). Retrieved from `http://dx.doi.org/10.21105/joss.00024` doi: 10.21105/joss.00024

Ganesh, T. V. (2016). cricketr: Analyze cricketers based on espn cricinfo statsguru [Computer software manual]. Retrieved from `http://CRAN.R-project.org/package=cricketr` (R package version 0.0.12)

Geman, S., & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*(6), 721–741.

Hamilton, J. D. (1994). *Time series analysis* (Vol. 2). Princeton University Press.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109.

ISO. (2012). *ISO/IEC 14882:2011 Information technology — Programming languages — C++*. Geneva, Switzerland: International Organization for Standardization. Retrieved from `http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=50372`

Jayadevan, V. (2002). A new method for the computation of target scores in interrupted, limited-over cricket matches. *Current Science*, *83*(5), 577–586.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, *106*(4), 620.

Jefferys, W. H., & Berger, J. O. (1992). Ockham's razor and bayesian analysis. *American Scientist*, *80*(1), 64–72.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*(282), 457–481.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Kimber, A. C., & Hansford, A. R. (1993). A statistical analysis of batting in cricket. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *153*(3), 443–455.

Koulis, T., Muthukumarana, S., & Briercliffe, C. D. (2014). A Bayesian stochastic model for batting performance evaluation in one-day cricket. *Journal of Quantitative Analysis in Sports*, *10*(1), 1–13.

Lang, D., Hogg, D. W., Jester, S., & Rix, H.-W. (2009). Measuring the undetectable: Proper motions and parallaxes of very faint sources. *The Astronomical Journal*, *137*(5), 4400.

Lemmer, H. H. (2004). A measure for the batting performance of cricket players. *South African Journal for Research in Sport, Physical Education and Recreation*, *26*(1), 55–64.

Lemmer, H. H. (2011). The single match approach to strike rate adjustments in batting performance measures in cricket. *Journal of Sports Science and Medicine*, *10*(4), 630.

MacKay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge University Press.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)*, *42*(2), 109–142.

Meyn, S. P., & Tweedie, R. L. (1993). Markov chains and stochastic stability. communication and control engineering series. *Springer-Verlag London Ltd., London*, *1*, 993.

Norman, J. M., & Clarke, S. R. (2010). Optimal batting orders in cricket. *Journal of the Operational Research Society*, *61*(6), 980–986.

O'Hagan, A., & Forster, J. J. (2004). *Kendall's advanced theory of statistics, volume 2b: Bayesian inference* (Vol. 2). Arnold.

Preston, I., & Thomas, J. (2000). Batting strategy in limited overs cricket. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *49*(1), 95–106.

R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Sivia, D., & Skilling, J. (2006). *Data analysis: a Bayesian tutorial*. Oxford University Press.

Skilling, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Analysis*, *1*(4), 833–859.

Stevenson, O. G., & Brewer, B. J. (2017). Bayesian survival analysis of opening batsmen in Test cricket. *Journal of Quantitative Analysis in Sports*, *13*(1), 25–36.

Swartz, T. B., Gill, P. S., Beaudoin, D., & de Silva, B. M. (2006). Optimal batting orders in one-day cricket. *Computers and Operations Research*, *33*(7), 1939–1950.

Swartz, T. B., Gill, P. S., & Muthukumarana, S. (2009). Modelling and simulation for one-day cricket. *Canadian Journal of Statistics*, *37*(2), 143–160.

Thomas, J. (2002). Rain rules for limited overs cricket and probabilities of victory. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *51*(2), 189–202.

Thorburn, W. M. (1918). The myth of Occam's razor. *Mind*, *27*(107), 345–353.

Totterdell, P. (1999). Mood scores: Mood and performance in professional cricketers. *British Journal of Psychology*, *90*, 317.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from `http://ggplot2.org`

Appendix A

New Zealand opening batsmen records and summaries

Table A.1: Test career batting records for all retired New Zealand opening batsmen since 1990. Players are listed in order of debut, from oldest to most recent.

Player	Matches	Innings	Not Outs	Runs	High-Score	Average	Strike Rate	100s	50s
D. White	2	4	0	31	18	7.75	33.33	0	0
B. Hartland	9	18	0	303	52	16.83	31.33	0	1
R. Latham	4	7	0	219	119	31.28	48.99	1	0
B. Pocock	15	29	0	665	85	22.93	29.80	0	6
B. Young	35	68	4	2034	267*	31.78	38.95	2	12
C. Spearman	19	37	2	922	112	26.34	41.68	1	3
M. Horne	35	65	2	1788	157	28.38	40.78	4	5
M. Bell	18	32	2	729	107	24.30	37.81	2	3
G. Stead	5	8	0	278	78	34.75	41.43	0	2
M. Richardson	38	65	3	2776	145	44.77	37.66	4	19
L. Vincent	23	40	1	1332	224	34.15	47.11	3	9
M. Papps	8	16	1	246	86	16.40	35.34	0	2
C. Cumming	11	19	2	441	74	25.94	34.86	0	1
J. Marshall	7	11	0	218	52	19.81	39.06	0	1
P. Fulton	23	39	1	967	136	25.44	39.27	2	5
J. How	19	35	1	772	92	22.70	50.45	0	4
A. Redmond	8	16	1	325	83	21.66	39.01	0	2
T. McIntosh	17	33	2	854	136	27.54	36.20	2	4
R. Nicol	2	4	0	28	19	7.00	26.66	0	0

Table A.2: Test career batting records for currently active New Zealand opening batsmen. Players are listed in order of debut, from oldest to most recent.

Player	Matches	Innings	Not Outs	Runs	High-Score	Average	Strike Rate	100s	50s
M. Guptill	47	89	1	2586	189	29.36	46.61	3	17
H. Rutherford	16	29	1	755	171	26.96	56.42	1	1
T. Latham	29	55	2	2140	177	40.37	46.59	6	12
J. Raval	4	8	1	237	55	33.85	47.59	0	2

Table A.3: Posterior summaries for all retired New Zealand opening batsmen since 1990, including the estimate for the next opener to debut for New Zealand. Players are listed in order of debut, from oldest to most recent.

Player	μ_1	**68% C.I.**	μ_2	**68% C.I.**	L	**68% C.I.**
D. White	$6.3^{+5.6}_{-3.3}$	[3.0, 11.9]	$16.7^{+13.9}_{-7.0}$	[9.7, 30.6]	$2.7^{+4.9}_{-2.1}$	[0.6, 7.6]
B. Hartland	$6.9^{+4.6}_{-2.9}$	[4.0, 11.5]	$20.7^{+6.6}_{-4.6}$	[16.1, 27.3]	$1.9^{+3.3}_{-1.4}$	[0.5, 5.2]
R. Latham	$10.5^{+9.9}_{-5.8}$	[4.7, 24.4]	$35.9^{+17.5}_{-10.6}$	[25.3, 53.4]	$4.1^{+7.6}_{-3.2}$	[0.9, 11.7]
B. Pocock	$8.4^{+5.5}_{-3.3}$	[5.1, 13.9]	$26.4^{+6.4}_{-4.7}$	[21.7, 32.8]	$1.9^{+3.4}_{-1.4}$	[0.5, 5.3]
B. Young	$15.0^{+5.9}_{-4.9}$	[10.1 , 20.1]	$36.0^{+6.4}_{-4.8}$	[31.2, 42.4]	$4.4^{+6.3}_{-3.0}$	[1.4, 10.7]
C. Spearman	$13.1^{+6.2}_{-4.8}$	[8.3, 19.3]	$28.8^{+5.9}_{-4.8}$	[24.0, 34.7]	$2.0^{+3.4}_{-1.5}$	[0.5, 5.4]
M. Horne	$14.3^{+5.4}_{-4.3}$	[10.0, 19.7]	$32.3^{+5.7}_{-4.5}$	[27.8, 38.0]	$4.4^{+5.1}_{-2.7}$	[1.7, 9.5]
M. Bell	$4.9^{+3.0}_{-1.8}$	[3.1, 7.9]	$32.7^{+10.1}_{-6.7}$	[26.0, 42.8]	$3.2^{+5.4}_{-2.5}$	[0.7, 8.6]
G. Stead	$18.0^{+11.5}_{-8.6}$	[9.4, 29.5]	$35.8^{+14.9}_{-9.8}$	[26.0, 50.7]	$3.1^{+6.7}_{-2.4}$	[0.7, 9.8]
M. Richardson	$30.7^{+8.5}_{-8.8}$	[21.9, 39.2]	$46.1^{+6.8}_{-5.6}$	[40.5, 52.9]	$3.6^{+6.9}_{-2.8}$	[0.8, 10.5]
L. Vincent	$13.5^{+7.0}_{-5.2}$	[8.3, 20.5]	$40.6^{+10.1}_{-7.4}$	[33.2, 50.7]	$6.0^{+7.7}_{-4.5}$	[1.5, 13.7]
M. Papps	$4.4^{+3.3}_{-1.9}$	[2.5, 7.7]	$23.6^{+11.0}_{-6.2}$	[17.4, 34.6]	$3.0^{+5.1}_{-2.2}$	[0.8, 8.1]
C. Cumming	$14.2^{+7.0}_{-5.7}$	[8.5, 21.2]	$29.3^{+9.3}_{-6.5}$	[22.8, 38.6]	$3.3^{+5.9}_{-2.5}$	[0.8, 9.2]
J. Marshall	$7.3^{+5.9}_{-3.7}$	[3.6, 13.2]	$24.0^{+9.8}_{-6.3}$	[17.7, 33.8]	$2.1^{+3.7}_{-1.6}$	[0.5, 5.8]
P. Fulton	$11.5^{+5.0}_{-4.0}$	[7.5, 16.5]	$31.1^{+8.4}_{-5.9}$	[25.2, 39.5]	$5.4^{+6.7}_{-3.4}$	[2.0, 12.1]
J. How	$10.1^{+5.6}_{-3.9}$	[6.2, 15.7]	$25.0^{+5.3}_{-4.1}$	[20.9, 30.3]	$1.2^{+2.7}_{-0.9}$	[0.3 3.9]
A. Redmond	$10.5^{+5.7}_{-4.2}$	[6.3, 16.2]	$28.2^{+11.9}_{-7.3}$	[20.9, 40.1]	$5.2^{+6.9}_{-3.4}$	[1.8, 12.1]
T. McIntosh	$8.0^{+4.5}_{-3.0}$	[5.0, 12.5]	$40.0^{+13.8}_{-9.2}$	[30.8, 53.8]	$9.2^{+8.0}_{-5.3}$	[3.9, 17.2]
R. Nicol	$5.9^{+5.4}_{-3.0}$	[2.9, 11.3]	$16.5^{+13.7}_{-7.0}$	[9.5, 30.2]	$3.0^{+5.0}_{-2.2}$	[0.8, 8.0]
NZ Opener	$10.1^{+11.7}_{-5.8}$	[4.3, 21.8]	$29.1^{+20.6}_{-12.1}$	[17.0, 49.7]	$3.2^{+5.9}_{-2.4}$	[0.8, 9.1]

Table A.4: Posterior summaries for currently active New Zealand opening batsmen, including the estimate for the next opener to debut for New Zealand. Players are listed in order of debut, from oldest to most recent.

Player	$\boldsymbol{\mu_1}$	**68% C.I.**	$\boldsymbol{\mu_2}$	**68% C.I.**	$\boldsymbol{L}$	**68% C.I.**
M. Guptill	$9.0^{+3.5}_{-2.6}$	[6.4, 12.5]	$34.1^{+5.0}_{-4.0}$	[34.0, 39.1]	$2.6^{+2.8}_{-1.4}$	[1.2, 5.4]
H. Rutherford	$15.3^{+6.8}_{-5.9}$	[9.4, 22.1]	$29.4^{+7.2}_{-5.3}$	[24.1, 36.6]	$2.3^{+6.2}_{-1.9}$	[0.4, 8.5]
T. Latham	$13.5^{+7.1}_{-4.7}$	[8.8, 20.6]	$46.9^{+8.9}_{-7.0}$	[39.9, 55.8]	$5.4^{+5.7}_{-3.5}$	[1.9, 11.1]
J. Raval	$16.7^{+11.8}_{-8.0}$	[8.7, 28.5]	$36.2^{+17.1}_{-10.8}$	[25.6, 53.3]	$3.7^{+6.7}_{-2.8}$	[1.1, 10.4]
NZ Opener	$10.1^{+11.7}_{-5.8}$	[4.3, 21.8]	$29.1^{+20.6}_{-12.1}$	[17.0, 49.7]	$3.2^{+5.9}_{-2.4}$	[0.8, 9.1]

Appendix B

International batsmen records and summaries

Table B.1: Test career batting records for international batsmen averaging 40+ since 2000 (30 innings minimum). Players are listed by country, in alphabetical order.

Player	Matches	Innings	Not Outs	Runs	High-Score	Average	Strike Rate	100s	50s
M. Clarke (AUS)	115	198	22	8643	329*	49.10	55.92	28	27
A. Gilchrist (AUS)	96	137	20	5570	204*	47.60	81.95	17	26
M. Hayden (AUS)	103	184	14	8625	380	50.73	60.10	30	59
M. Hussey (AUS)	79	137	16	6235	195	51.52	50.13	19	29
S. Katich (AUS)	56	99	6	4188	157	45.03	49.36	10	25
J. Langer (AUS)	105	182	12	7696	250	45.27	54.22	23	30
D. Lehmann (AUS)	27	42	2	1798	177	44.95	61.80	5	10
D. Martyn (AUS)	67	109	14	4406	165	46.37	51.41	13	23
R. Ponting (AUS)	168	287	29	13378	257	51.85	58.72	41	62
C. Rogers (AUS)	25	48	1	2015	173	42.87	50.60	5	14
A. Symonds (AUS)	26	41	5	1462	162*	40.61	64.80	2	10
M. Waugh (AUS)	128	209	17	8029	153*	41.81	52.27	20	47
S. Waugh (AUS)	168	260	46	10927	200	51.06	48.64	32	50

Player	Matches	Innings	Not Outs	Runs	High-Score	Average	Strike Rate	100s	50s
I. Bell (ENG)	118	205	24	7727	235	42.69	49.46	22	46
K. Pietersen (ENG)	104	181	8	8181	227	47.28	61.72	23	35
M. Prior (ENG)	79	123	21	4099	131*	40.18	61.66	7	28
A. Strauss (ENG)	100	178	6	7037	177	40.91	48.91	21	27
G. Thorpe (ENG)	100	179	28	6744	200*	44.66	45.89	16	39
M. Trescothick (ENG)	76	143	10	5825	219	43.79	54.51	14	29
J. Trott (ENG)	52	93	6	3835	226	44.08	47.18	9	19
M. Vaughan (ENG)	82	147	9	5719	197	41.44	51.13	18	18
R. Dravid (IND)	164	286	32	13288	270	52.31	42.51	36	63
S. Ganguly (IND)	113	188	17	7212	239	42.17	51.25	16	35
V. Laxman (IND)	134	225	34	8781	281	45.97	49.37	17	56
V. Sehwag (IND)	104	180	6	8586	319	49.34	82.23	23	32
S. Tendulkar (IND)	200	329	33	15921	248*	53.78	N/A	51	68
S. Fleming (NZ)	111	189	10	7172	274*	40.06	45.82	9	46
M. Richardson (NZ)	38	65	3	2776	145	44.77	37.66	4	19
J. Ryder (NZ)	18	33	2	1269	201	40.93	55.19	3	6
I. ul-Haq (PAK)	120	200	22	8830	329	49.60	54.02	25	46
M. Yousuf (PAK)	90	156	12	7530	223	52.29	52.39	24	33
H. Gibbs (SA)	90	154	7	6167	228	41.95	50.26	14	26
J. Kallis (SA)	166	280	40	13289	224	55.37	45.97	45	58
G. Kirsten (SA)	101	176	15	7289	275	45.27	43.43	21	24
A. Prince (SA)	66	104	16	3665	162*	41.64	43.70	11	11
G. Smith (SA)	117	205	13	9265	277	48.25	59.67	27	38
T. Dilshan (SL)	87	145	11	5492	193	40.98	65.54	16	23
S. Jayasuriya (SL)	110	188	14	6973	340	40.07	N/A	14	31
M. Jayawardene (SL)	149	252	15	11814	374	49.84	51.45	34	50
H. Tillakaratne (SL)	83	131	25	4545	204*	42.87	N/A	11	20
T. Samaraweera (SL)	81	132	20	5462	231	48.76	46.92	14	30
K. Sangakkara (SL)	134	233	17	12400	319	57.40	54.19	38	52
S. Chanderpaul (WI)	164	280	49	11867	203*	51.37	43.31	30	66
C. Gayle (WI)	103	182	11	7214	333	42.18	60.26	15	37
B. Lara (WI)	131	232	6	11953	400*	52.88	60.51	34	48
A. Flower (ZIM)	63	112	19	4794	232*	51.54	45.07	12	27

Table B.2: Model marginal likelihoods or 'evidence' for the exponential varying-hazard model, Gaussian hazard model and AR(1) hazard model.

Player	**Exponential varying-hazard**	**Gaussian hazard**	**AR(1) hazard**
M. Clarke (AUS)	−824.14	−824.03	−824.24
A. Gilchrist (AUS)	−543.79	−543.91	−544.18
M. Hayden (AUS)	−815.47	−814.92	−814.58
M. Hussey (AUS)	−582.87	−582.59	−583.16
S. Katich (AUS)	−430.58	−430.47	−430.71
J. Langer (AUS)	−800.83	−800.86	−801.17
D. Lehmann (AUS)	−194.63	−194.66	−194.92
D. Martyn (AUS)	−442.86	−442.80	−443.10
R. Ponting (AUS)	−1236.14	−1236.24	−1236.34
C. Rogers (AUS)	−209.58	−209.32	−209.23
A. Symonds (AUS)	−170.03	−170.19	−170.31
M. Waugh (AUS)	−887.51	−887.26	−887.47
S. Waugh (AUS)	−1032.36	−1032.15	−1032.14
I. Bell (ENG)	−806.93	−806.84	−806.92
K. Pietersen (ENG)	−816.35	−816.31	−816.55
M. Prior (ENG)	−455.65	−455.40	−455.11
A. Strauss (ENG)	−782.03	−781.83	−782.22
G. Thorpe (ENG)	−725.23	−725.28	−725.29
M. Trescothick (ENG)	−634.50	−634.63	−635.06
J. Trott (ENG)	−389.77	−389.70	−390.02
M. Vaughan (ENG)	−654.75	−654.59	−694.97
R. Dravid (IND)	−1262.20	−1262.44	−1262.87
S. Ganguly (IND)	−811.08	−811.25	−811.45
V. Laxman (IND)	−921.71	−921.82	−921.92
V. Sehwag (IND)	−847.51	−847.50	−847.88
S. Tendulkar (IND)	−1475.78	−1475.94	−1475.44
S. Fleming (NZ)	−836.19	−836.22	−836.54
M. Richardson (NZ)	−302.39	−302.25	−302.45
J. Ryder (NZ)	−146.16	−146.21	−146.27
I. ul-Haq (PAK)	−848.55	−848.63	−848.77
M. Yousuf (PAK)	−691.83	−691.88	−692.31
H. Gibbs (SA)	−696.48	−696.92	−696.82
J. Kallis (SA)	−1172.66	−1172.78	−1173.42

Player	Exponential varying-hazard	Gaussian hazard	AR(1) hazard
G. Kirsten (SA)	−769.79	−770.04	−770.18
A. Prince (SA)	−405.67	−405.55	−405.69
G. Smith (SA)	−906.77	−906.88	−907.28
T. Dilshan (SL)	−603.04	−603.07	−603.34
S. Jayasuriya (SL)	−815.59	−815.68	−815.76
M. Jayawardene (SL)	−1104.75	−1104.86	−1104.93
H. Tillakaratne (SL)	−499.67	−499.59	−499.62
T. Samaraweera (SL)	−544.04	−544.21	−544.46
K. Sangakkara (SL)	−1034.86	−1034.65	−1034.94
S. Chanderpaul (WI)	−1122.55	−1122.84	−1123.17
C. Gayle (WI)	−790.65	−790.77	−790.94
B. Lara (WI)	−1114.95	−1115.17	−1115.33
A. Flower (ZIM)	−462.33	−462.39	−462.32

Predictive hazard functions

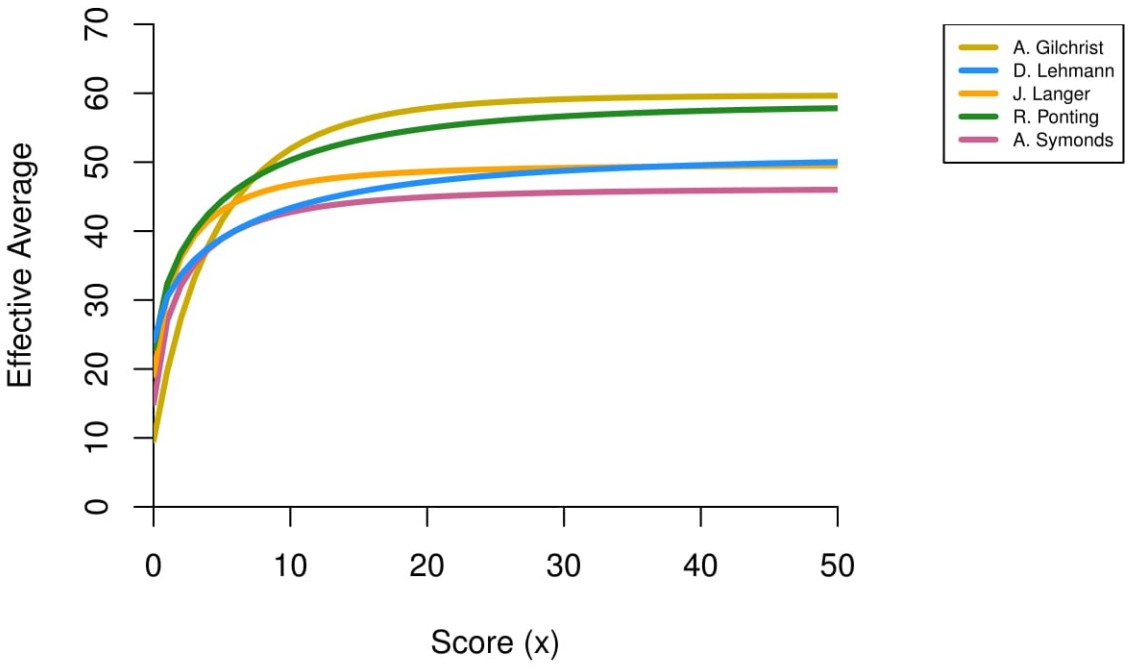


Figure B.1. Predictive hazard functions for the exponential varying-hazard model in terms of effective average, $\mu(x)$, for Australian batsmen.

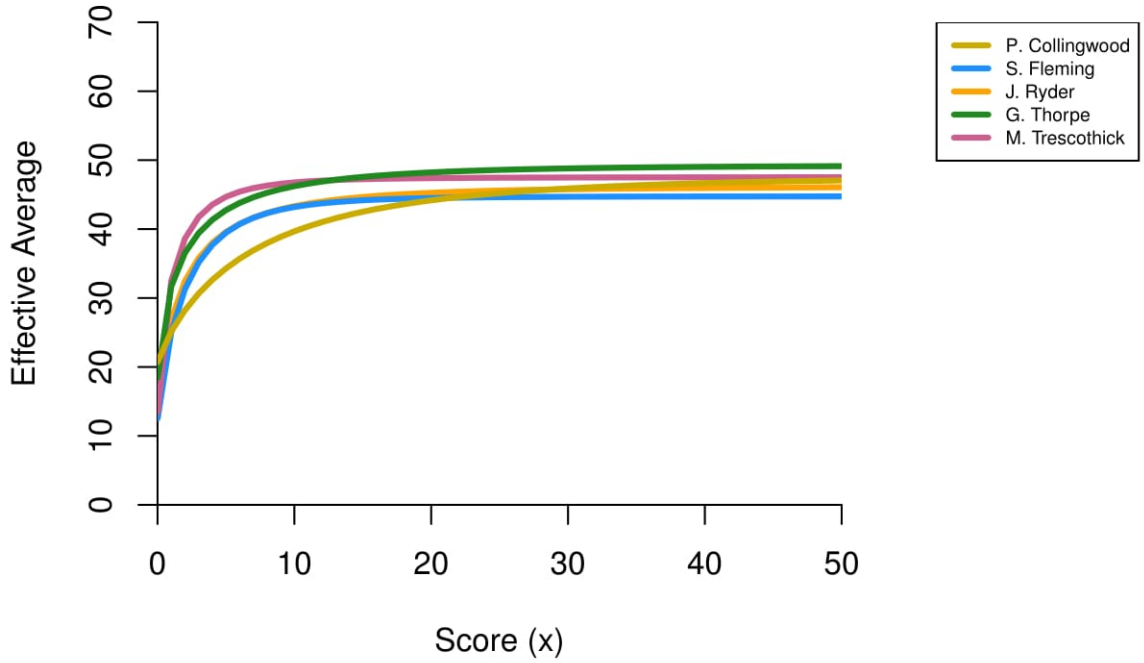


Figure B.2. Predictive hazard functions for the exponential varying-hazard model in terms of effective average, $\mu(x)$, for English and New Zealand batsmen.

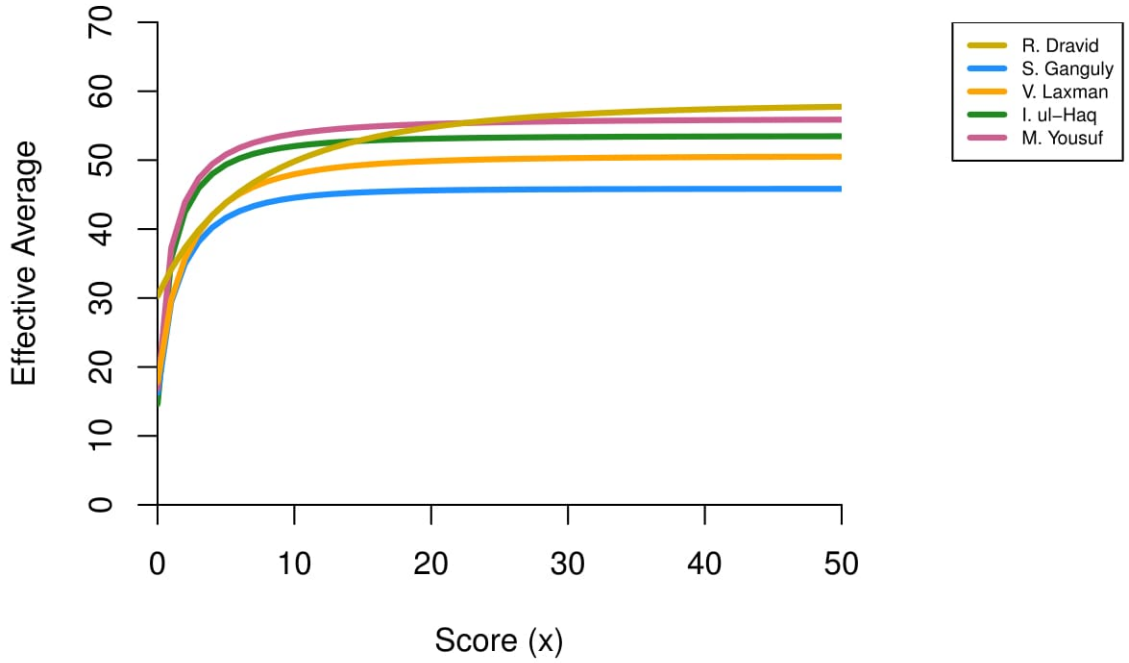


Figure B.3. Predictive hazard functions for the exponential varying-hazard model in terms of effective average, $\mu(x)$, for Indian and Pakistani batsmen.

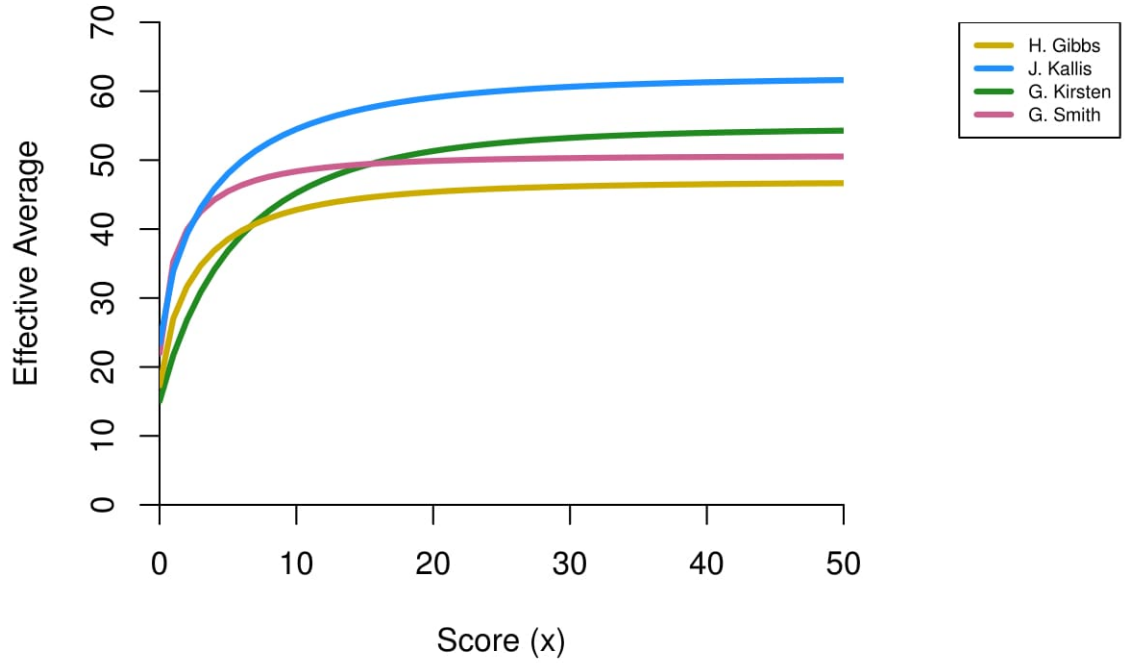


Figure B.4. Predictive hazard functions for the exponential varying-hazard model in terms of effective average, $\mu(x)$, for South African batsmen.

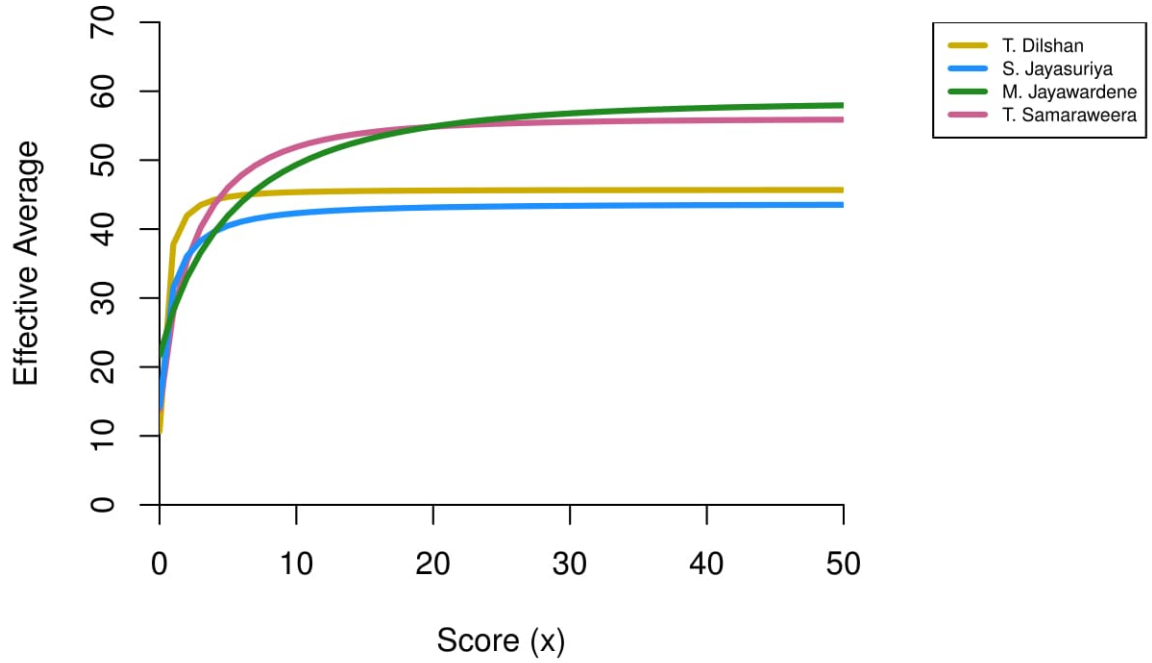


Figure B.5. Predictive hazard functions for the exponential varying-hazard model in terms of effective average, $\mu(x)$, for Sri Lankan batsmen.

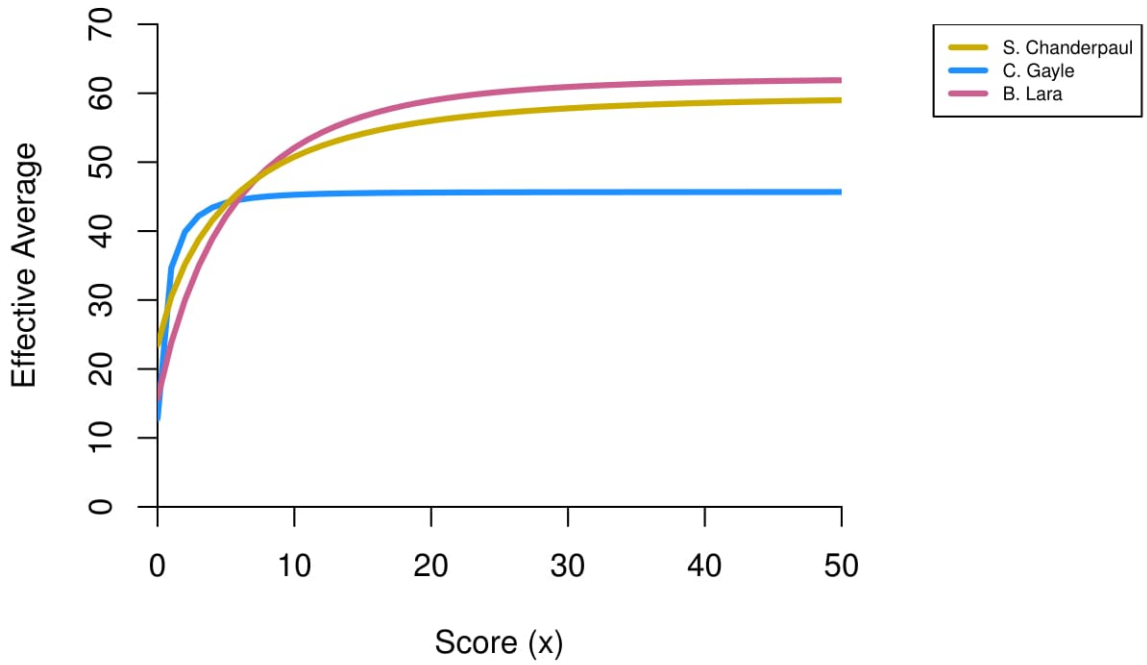


Figure B.6. Predictive hazard functions for the exponential varying-hazard model in terms of effective average, $\mu(x)$, for West Indian batsmen.

Appendix C

Model code and data

Exponential varying-hazard model code

The following Julia functions were used to implement the exponential varying-hazard model. Each function was fed into the nested sampling algorithm accordingly to produce posterior samples for a given batsman.

```
@doc """
An object of this class represents a point in parameter space.
There are functions defined to evaluate the log likelihood and
move around.
""" ->
type Particle
    params::Vector{Float64}
end


@doc """
A constructor. We have 6 parameters:
The usual mu_1, mu_2 and the time scale L
As well as the Gaussian parameters k, sigma and m (strength, width and midpoint of Gaussian)
""" ->
function Particle()
    return Particle(Array(Float64, (6, )))
end


@doc """
Generate parameters from the prior
""" ->
function from_prior!(particle::Particle)
    ## mu2 ~ Lognormal(25, 0.75)
    particle.params[2] = rand(Normal(log(25), 0.75), 1)[1]

    ## C ~ Beta(1, 2), D ~ Beta(1, 5)
    particle.params[1] = rand(Beta(1, 2), 1)[1]
    particle.params[3] = rand(Beta(1, 5), 1)[1]

    return nothing
end
```

```
@doc """
Do a metropolis proposal. Return log(hastings factor for prior sampling)
""" ->
function perturb!(particle::Particle)
    ## Define logH, length of data (i.e. number of innings played by this batsman)
    logH = 0.0
    innings = length(data[:, 1])

    ## Randomly decide which parameter we are going to evolve
    i = rand(1:length(particle.params))

    ## Evolve C
    if(i == 1)

        ## Beta prior for C
        a = 1
        b = 2

        ## Log-prior before
        logH -= (a - 1) * log(particle.params[i]) + (b - 1) * log(1 - particle.params[i])

        ## Explore parameter space - map it onto the [0, 1] interval
        particle.params[i] += 1 * randh()
        particle.params[i] = mod(particle.params[i], 1)

        ## Log-prior after
        logH += (a - 1) * log(particle.params[i]) + (b - 1) * log(1 - particle.params[i])
    end

    ## Evolve mu2
    if(i == 2)

        ## Lognormal prior for mu2
        mu = log(25)
        sig = 0.75

        ## Log-prior before
        logH -= -0.5 * log(2 * pi) - log(sig) - (1/(2 * sig^2)) * (particle.params[i] - mu)^2

        ## Explore the parameter space (use a scale of sigma)
        particle.params[i] = particle.params[i] + sig * randh()

        ## Log-prior after
        logH += -0.5 * log(2 * pi) - log(sig) - (1/(2 * sig^2)) * (particle.params[i] - mu)^2
    end

    ## Evolve D
    if(i == 3)

        ## Beta prior
        a = 1
        b = 5

        ## Log-prior before
        logH -= (a - 1) * log(particle.params[i]) + (b - 1) * log(1 - particle.params[i])

        ## Explore parameter space - map it onto the [0, 1] interval
        particle.params[i] += 1 * randh()
        particle.params[i] = mod(particle.params[i], 1)

        ## Log-prior after
        logH += (a - 1) * log(particle.params[i]) + (b - 1) * log(1 - particle.params[i])
    end

    ## Return the difference in likelihoods before and after
    return logH
end
```

```
@doc """
The 'effective average' function
"""->
## Using mu_1 = C * mu_2, L = D * mu_2
function effective_average(particle::Particle, data::Array{Float64} = data)
    ## Take the exponential of mu2, exp(Normal(log(25), 0.75)) == Lognormal(25, 0.75)
    mu2 = exp(particle.params[2])
    ## mu1 = C * mu2
    mu1 = particle.params[1] * mu2
    ## L = D * mu2
    L = particle.params[3] * mu2
    return(mu2 + (mu1 - mu2) * exp(-data[:, 1] / L))
end


@doc """
Evaluate the log likelihood
""" ->
function log_likelihood(particle::Particle, data::Array{Float64, 2} = data)
    logL1 = 0.0
    logL2 = 0.0

    ## Vector of scores from 0 to batsman's maximum score
    scores = Float64[0:maximum(data[:, 1]); ]

    ## In terms of the hazard function, H(x)
    cumsum_scores = cumsum(log(effective_average(particle, scores)) - log(effective_average(particle, scores) + 1))

    ## Take the x_i - 1 cumulative sum for each score
    for(i in 1:length(data[:, 1]))
        score = Int64[data[i, 1]][1] # convert to integer for indexing (cannot index using Float64)
        if(score >= 1)
            logL1 += cumsum_scores[score]
        end
    end

    ## Out scores
    out = data[data[:, 2] .== 0, :]
    logL2 = Float64[sum(-log(effective_average(particle, out[:, 1]) + 1))]

    ## Return the log-likelihood
    return(logL1[1] + logL2[1])
end
```

Gaussian hazard model code

In order to implement the Gaussian hazard model, changes are only required for functions `Particle()`, `from_prior()`, `perturb()` and `effective_average()`. The `log_likelihood()` function remains the same.

```
@doc """
A constructor. We have 6 parameters:
The C, mu2 and D from the exponential varying-hazard model, as well as the Gaussian parameters
k, phi and m (strength, width and midpoint of Gaussian)
""" ->
function Particle()
    return Particle(Array(Float64, (6, )))
end
```

```
@doc """
Generate parameters from the prior
""" ->
function from_prior!(particle::Particle)
    ## mu2 ~ Lognormal(25, 0.75)
    particle.params[2] = rand(Normal(log(25), 0.75), 1)[1]

    ## C ~ Beta(1, 2), D ~ Beta(1, 5)
    particle.params[1] = rand(Beta(1, 2), 1)[1]
    particle.params[3] = rand(Beta(1, 5), 1)[1]

    ## k ~ Uniform(-1, 1)
    particle.params[4] = rand(Uniform(-1, 1), 1)[1]

    ## phi ~ Uniform(0, 20)
    particle.params[5] = rand(Uniform(0, 20), 1)[1]

    ## midpoint ~ Uniform(0, 400)
    particle.params[6] = rand(Uniform(0, 400), 1)[1]

    return nothing
end


@doc """
Do a metropolis proposal. Return log(hastings factor for prior sampling)
""" ->
function perturb!(particle::Particle)
    ## Define logH, length of data (i.e. number of innings played by this batsman)
    logH = 0.0
    innings = length(data[:, 1])

    ## Randomly decide which parameter we are going to evolve
    i = rand(1:length(particle.params))

    ## Evolve C
    if(i == 1)

        ## Beta prior for C
        a = 1
        b = 2

        ## Log-prior before
        logH -= (a - 1) * log(particle.params[i]) + (b - 1) * log(1 - particle.params[i])

        ## Explore parameter space - map it onto the [0, 1] interval
        particle.params[i] += 1 * randh()
        particle.params[i] = mod(particle.params[i], 1)

        ## Log-prior after
        logH += (a - 1) * log(particle.params[i]) + (b - 1) * log(1 - particle.params[i])
    end

    ## Evolve mu2
    if(i == 2)

        ## Lognormal prior for mu2
        mu = log(25)
        sig = 0.75

        ## Log-prior before
        logH -= -0.5 * log(2 * pi) - log(sig) - (1/(2 * sig^2)) * (particle.params[i] - mu)^2

        ## Explore the parameter space (use a scale of sigma)
        particle.params[i] = particle.params[i] + sig * randh()

        ## Log-prior after
        logH += -0.5 * log(2 * pi) - log(sig) - (1/(2 * sig^2)) * (particle.params[i] - mu)^2
    end
```

```
    ## Evolve D
    if(i == 3)

        ## Beta prior
        a = 1
        b = 5

        ## Log-prior before
        logH -= (a - 1) * log(particle.params[i]) + (b - 1) * log(1 - particle.params[i])

        ## Explore parameter space - map it onto the [0, 1] interval
        particle.params[i] += 1 * randh()
        particle.params[i] = mod(particle.params[i], 1)

        ## Log-prior after
        logH += (a - 1) * log(particle.params[i]) + (b - 1) * log(1 - particle.params[i])
    end

    ## Evolve k
    if(i == 4)
        ## Standard deviation
        sig = 2/sqrt(12)

        ## Explore parameter space - map it onto the [-1, 1] interval
        particle.params[i] += sig * randh()
        negpos = sample([-1, 1])
        particle.params[i] = mod(particle.params[i], 1) * negpos
    end

    ## Evolve phi
    if(i == 5)
        ## Standard deviation
        sig = 20/sqrt(12)

        ## Explore parameter space - map it onto the [0, 20] interval
        particle.params[i] += sig * randh()
        particle.params[i] = mod(particle.params[i], 20)
    end

    ## Alter midpoint of Gaussian
    if(i == 6)
        ## Standard deviation
        sig = 400/sqrt(12)

        ## Explore parameter space - map it onto the [0, 400] interval
        particle.params[i] += sig * randh()
        particle.params[i] = mod(particle.params[i], 400)
    end

    ## Return the difference in likelihoods before and after
    return logH
end

@doc """
The 'effective average' function
"""->
## Using mu_1 = C * mu_2, L = D * mu_2
function effective_average(particle::Particle, data::Array{Float64} = data)
    ## Take the exponential of mu2, exp(Normal(log(25), 0.75)) == Lognormal(25, 0.75)
    mu2 = exp(particle.params[2])
    ## mu1 = C * mu2
    mu1 = particle.params[1] * mu2
    ## L = D * mu2
    L = particle.params[3] * mu2

    ## Gaussian components
    k = particle.params[4]
    phi = particle.params[5]
    m = particle.params[6]

    ## Construct the gaussian function
    gaussian = exp(-k * exp(-1/(2 * phi^2) * (data[:, 1] - m).^2))

    ## Multiple the underlying effective average by the Gaussian function
    return(mu2 + (mu1 - mu2) * exp(-data[:, 1] / L))
end
```

AR(1) hazard model code

As the AR(1) hazard model uses a C++ implementation of the diffusive nested sampling algorithm, the only function written in Julia was the `effective_average()` function that is used to evaluate the likelihood function (the likelihood function remains the same from the exponential varying-hazard and Gaussian hazard models).

It is worth noting that the C++ implementation uses a Uniform(0, 1) prior for all parameters. Therefore, to obtain the actual parameter values, we must use inverse transform sampling, by using the uniform distributed value as the quantile value for the actual prior distribution, for each respective parameter.

```
@doc """
The 'effective average' function.
Since the C++ implementation of diffusive nested sampling samples all parameters from
a Uniform(0, 1) distribution, we must use inverse transform sampling to get the actual
parameter values.
This can be done by taking the quantile of the respective prior distribution for each parameter,
using the uniformly distributed number generated in C++.
"""->
function effective_average(particle::Particle, data::Array{Float64} = data)
    ## Take the exponential of mu2, exp(Normal(log(25), 0.75)) == Lognormal(25, 0.75)
    mu2 = exp(quantile(Normal(log(25), 0.75), params[2]))
    ## mu1 = C * mu2
    mu1 = quantile(Beta(1, 2), params[1]) * mu2
    ## L = D * mu2
    L = quantile(Beta(1, 5), params[3]) * mu2

    ## AR(1) parameter values
    s = Array{Float64}(401)
    alpha = quantile(Beta(4, 1), params[4])
    beta = quantile(Exponential(0.1), params[5])
    noise = quantile(Normal(0, 1), params[6:406]) # 400 noise parameters

    ## Construct the AR(1) process
    sigma = beta/sqrt(1.0 - alpha^2) # standard deviation
    s[1] = noise[1] * sigma
    for(i in 2:401)
        s[i] = alpha * y[i - 1] + noise[i] * beta
    end

    ## Return the underlying effective average multiplied by the AR(1) process
    return(mu2 + (mu1 - mu2) * exp(-data[:, 1] / L) .* exp(s))
end
```

Data

Australia

Michael Clarke

151, 17, 5, 39*, 91, 73, 17, 7, 141, 7, 1, 27, 20, 35, 8, 8, 22, 11, 91, 40, 30, 7, 39, 36, 56, 25, 39, 5, 5, 14*, 5, 19, 9, 23*, 56, 124, 21*, 37, 135*, 5, 11, 145*, 71, 20, 73, 1, 0, 23, 81, 118, 110, 10, 0, 48*, 11, 6, 23, 69, 112, 8, 22, 98, 9, 110, 62, 25, 88*, 29, 138, 41, 68, 0, 3, 23*, 0, 47, 83, 1, 136, 29, 103*, 93, 3, 0, 41, 71, 61*, 11, 25, 28*, 37, 3, 21, 166, 168, 28, 63, 14, 4, 14, 3, 9, 2, 80, 4, 20, 20, 13, 4, 41, 23, 60, 13, 6, 112, 151, 2, 11, 2, 139, 22, 0, 31, 1, 329*, 18, 210, 37, 73, 6, 45, 15, 24, 25, 259*, 230, 38, 5, 44, 74, 57*, 106, 50, 29, 130, 31, 91, 16, 0, 18, 0, 23, 28, 51, 187, 30*, 6, 21, 7, 28*, 1, 113, 148, 22, 24, 23, 10, 6*, 10, 6, 23, 17*, 19, 1, 161*, 0, 128, 7, 18, 47, 14*, 38, 4, 7, 32*, 10, 3, 10, 13, 15

Adam Gilchrist

81, 6, 149*, 28, 0, 43, 78, 55, 45*, 7, 59, 3, 75, 0*, 48, 50, 9, 10*, 37, 87, 122, 0, 0, 1, 1, 152, 90, 54, 19, 25, 118, 20, 39, 0, 83*, 7, 22, 30*, 34, 204*, 138*, 24, 91, 16, 0, 60*, 54, 38, 1, 10*, 133, 37, 77, 101*, 65, 33, 6, 43, 113*, 20, 0, 29, 43, 14, 6, 4, 4, 0, 0, 144, 22, 31*, 0, 80, 35, 0, 104, 26, 3, 49, 2, 3*, 26, 5, 126, 50, 69, 0*, 48, 113, 121, 162, 60*, 26, 10, 49*, 1, 30, 4, 27, 11, 23, 94, 1, 44, 2, 6, 6, 44, 2, 0, 86, 12, 2, 24, 12, 0, 144, 12, 0, 64, 0, 102*, 1, 62, 67*, 23, 35, 7, 1, 55, 15, 14

Matthew Hayden

15, 5, 5, 0, 125, 0, 47, 40, 0, 14, 10, 0, 2, 37, 44, 69, 58, 14, 13, 30, 3, 5, 119, 28*, 97, 67, 203, 35, 35, 0, 6*, 33, 42, 15, 35, 68, 136, 13, 91, 0, 57, 31, 131, 138, 3*, 105, 21*, 122, 63, 96, 28, 0, 197, 103, 46, 30, 102, 1, 15, 2, 10, 19, 30, 100*, 27, 2*, 14, 177, 11, 50, 380, 20, 101*, 37, 99, 12, 17, 136, 53*, 67, 30, 41, 130, 54, 5, 25, 28, 37, 2, 117, 132, 26, 30, 58, 39, 23, 9, 35, 24, 8, 70, 54, 4, 10, 9, 56*, 26, 23*, 35, 15, 61, 38, 9, 12,

34, 0, 31, 34, 36, 7, 26, 138, 0*, 111, 77, 37, 118, 110, 46, 47, 87*, 0, 20, 65, 137, 4, 90, 94, 32, 0, 102, 3, 0, 6, 72, 29, 21, 37, 12, 18, 24, 92, 153, 33, 23*, 43, 17, 33, 124, 47, 13, 123, 103, 0, 13, 0, 29, 83, 16*, 16, 77, 8, 0, 24, 12, 4, 8, 23, 31, 39

Michael Hussey

1, 29, 137, 31*, 133*, 30*, 23, 58, 122, 31, 45, 6, 14*, 75, 73, 89, 23, 37, 182, 86, 91, 61*, 74*, 103, 6, 37, 133, 132, 34*, 2, 36, 41, 145*, 0, 46, 22, 56, 1, 10, 40, 12, 18, 146, 31, 54, 1, 53, 90, 19, 35, 0, 70, 0, 8, 0, 2, 30, 45*, 4, 0, 50, 19, 20, 39, 3, 51, 27, 0, 64, 10, 0, 121, 66, 41, 29, 82, 17, 82, 4, 28, 134*, 6, 13*, 4, 22, 67, 17, 28, 34, 20, 195, 93, 52, 61, 116, 8, 0, 33, 12, 95, 15, 142, 118, 93, 1, 0, 20, 39, 15, 8, 0, 0, 89, 150*, 14, 25, 15, 48, 32, 73, 24, 10, 32, 100, 103, 54, 12, 26, 115*, 31*, 34, 25, 27*

Simon Katich

15, 0*, 52, 16, 75, 31, 29, 125, 77*, 14, 86, 9, 15, 1, 1, 81, 39, 36*, 9, 4, 99, 7, 1, 118, 35, 35, 27, 67, 4, 16, 17, 12, 45, 59, 1, 0, 2, 0, 12, 1, 113, 36, 157, 66, 34, 33, 20, 64, 14*, 102, 16, 10, 131*, 23, 83, 37, 54, 15, 47, 61, 3, 10, 108, 30, 55, 54, 122, 48, 6, 46, 26, 0, 50, 43, 92, 80, 21, 99, 10, 98, 2, 11, 100, 79, 18*, 88, 106, 6, 37, 43, 24, 50, 4, 0, 43

Justin Langer

20, 54, 10, 1, 63, 24, 0, 0, 69, 12, 0, 19, 0, 116, 14, 30, 51, 8, 74, 15, 7, 179*, 52, 44, 30, 26, 1, 5, 24, 8, 24, 51, 1, 51, 127, 7, 5, 7, 32, 44, 1, 59, 127, 144, 11, 38, 8, 9, 223, 46, 47, 12, 57, 4, 122*, 3, 5, 6, 48, 31, 80, 20, 10, 19, 58, 28, 35, 21, 102*, 104, 18*, 123, 75, 0, 116, 1, 85, 7, 126, 30*, 28, 37, 58, 11, 18, 32, 22, 48, 19, 250, 24, 25, 3, 146, 78*, 25, 3, 78, 0, 42, 111, 71, 1, 26, 2, 8, 121, 0, 58, 10, 14, 2, 117, 47, 12, 32, 3, 9, 19, 166, 30, 10, 162, 8, 52, 0, 71, 19, 44, 30, 12, 0, 34, 215, 46, 191, 97, 50, 5, 13, 34, 23, 72*, 46, 6, 59*, 40, 6, 82, 28, 31, 14, 27, 61, 105, 0*, 0, 22, 99, 20, 37, 47, 25, 20, 16, 34, 35, 37, 0*, 82, 100*, 4, 7, 37, 0, 27, 26, 20*

Darren Lehmann

52, 98, 3, 26, 13, 4, 32, 0, 30, 20*, 5, 42, 6, 160, 66, 96, 4*, 7, 14, 110, 177, 30, 63, 129, 8, 21, 153, 1, 57, 51, 50, 21, 17, 14, 0, 31, 70, 8, 81, 12, 5, 11

Damien Martyn

36, 15, 7, 67*, 0, 13, 31, 1, 74, 8, 59, 6, 17, 36, 78, 17*, 89*, 4, 46*, 34*, 105, 52, 4, 33*, 118, 6, 64*, 4, 0, 60, 30, 124*, 6*, 52, 117, 133, 2, 0, 11, 0, 26, 64, 95, 71, 17, 0, 26, 21, 53, 32, 42, 66*, 30, 38, 31, 7, 40, 42, 110, 1, 161, 14, 5, 47, 7, 97, 52, 3, 45, 26, 104, 114, 97, 55, 0, 70, 7, 6*, 1, 100*, 142, 67, 32, 165, 38, 2, 65, 20, 28, 20, 19, 1, 13, 10, 22, 9*, 57, 15*, 21, 101, 4, 7, 29, 11, 5

Ricky Ponting

96, 71, 6, 20, 14, 13, 88, 9, 9, 4, 127, 9, 45, 40, 20, 26, 73*, 16, 4, 105, 32, 62, 26, 23, 18, 2, 60, 9, 16, 76*, 43, 21, 11, 5, 10, 104, 22, 21, 21*, 96, 51, 1, 105*, 31, 0, 0, 0, 197, 125, 21, 67, 21*, 141*, 20, 5, 92, 11, 23, 26*, 51, 14*, 0, 6, 0, 0, 11, 11, 14, 4, 14, 17, 144, 72, 62, 5, 32*, 157*, 31, 26, 54, 25, 22, 0*, 14, 39, 47, 100*, 89, 34, 123, 3, 154, 68, 21, 30, 7, 11, 117, 42*, 206, 45, 113, 10, 59, 37, 169, 53*, 54, 50, 242, 0, 257, 31*, 25, 47, 21, 28, 10, 27, 92, 20, 22, 45, 11, 12, 51, 68, 26*, 25, 98, 7, 62*, 207, 4*, 46, 47*, 9, 105, 86*, 9, 42, 61, 0, 7, 156, 1, 48, 35, 46, 54, 149, 104*, 17, 0*, 56, 3, 71, 53, 117, 11, 120, 143*, 74, 1, 103, 116, 34, 20, 21, 118*, 52, 196, 60*, 142, 49, 2, 75, 7, 45, 56, 31, 53*, 4, 3, 55, 1, 20, 45, 140, 158, 5, 65, 38, 18, 39, 123, 17, 5, 2, 87, 24, 8, 4, 17, 79, 0, 32, 101, 99, 0, 53, 83, 25, 9, 81, 0, 12, 150, 2, 38, 38, 5, 78, 8, 66, 55, 36, 20, 23*, 2, 57, 12, 0, 11, 209, 89, 41, 22, 6, 71, 4, 77, 72, 10, 51*, 0, 9, 12, 1, 10, 20, 44, 4, 48, 28, 8, 0, 0, 62, 78, 5, 16, 62, 60, 134, 7, 221, 60*, 4, 14, 7, 41, 23, 57, 0, 4, 16, 4, 8

Chris Rogers

4, 15, 16, 52, 15, 6, 84, 12, 110, 49, 23, 1, 16, 72, 2, 11, 54, 61, 116, 11, 119, 4, 1, 5, 107, 25, 39, 9, 21, 55, 55, 57, 69, 95, 56, 95, 10, 173, 49*, 52, 6, 0, 52, 43

Andrew Symonds

0, 24, 6, 23, 1, 9, 13, 25, 0, 72, 12, 55, 13, 4, 29, 26, 2, 156, 48, 53*, 50*, 35, 44, 162*, 61, 66, 12, 30, 70*, 79, 18, 43*, 52, 2, 26, 20, 0, 57, 37, 27, 0

Mark Waugh

138, 23, 26, 39, 71, 31, 64, 20*, 3, 139*, 0, 11, 34, 5, 18, 15, 0, 5, 56, 0, 0, 0, 0, 39, 60, 112, 16, 57, 0, 26, 9, 21, 13, 12, 6, 64, 99, 70, 1, 52, 137, 62*, 10, 49, 36, 111, 68, 84, 7, 11, 2, 12, 42, 28, 7, 43, 113*, 20, 61, 68, 71, 140, 15, 71, 29, 3, 25, 39, 24, 88, 1, 40, 4, 61, 2, 7, 126, 59, 88, 3, 116, 34, 111, 61, 71, 12, 26, 23, 38, 57, 19, 67, 0, 19, 82, 79, 9, 26, 20, 116, 5, 42, 5, 1, 33, 12, 55, 8, 68, 7, 19, 1, 3, 17, 86, 81, 9, 0, 1, 100, 63, 115*, 66, 18, 10, 0, 153*, 33*, 0, 42, 43, 26, 117, 31, 27*, 36, 17*, 7, 51*, 36, 43, 121, 24, 2, 33, 67, 21, 0, 3, 11, 65, 6, 0, 10, 13, 90, 100, 5, 0, 0, 5, 8, 41, 51*, 32, 72*, 25, 3, 44*, 28, 18, 24, 119, 63, 5, 25, 78*, 22, 3, 0, 22, 0, 70, 57, 49, 108, 0*, 15, 42*, 72, 24*, 120, 0, 12, 42, 86, 2, 74, 34, 19, 53, 25, 16, 45, 30

Steve Waugh

13, 5, 8, 0, 11, 74, 1, 1, 0, 39*, 6, 0, 28, 71, 79*, 10, 49, 0, 73, 21, 61, 55, 10, 27, 20, 0, 13, 1, 19, 59, 4, 90, 91, 26, 42, 3, 55*, 12, 8, 177*, 152*, 21*, 43, 92, 0, 14, 7*, 17, 60, 57, 16, 134*, 20, 3, 17, 4, 25, 25, 1, 19, 48, 14, 26, 2, 4*, 10, 20, 38, 1, 100, 42, 4, 13, 0, 62, 75, 41, 0, 3, 78*, 13*, 13, 47*, 157*, 59, 20, 26, 44, 25*, 147*, 164, 1, 45*, 0, 86, 64, 73, 0, 98, 19, 7, 94*, 26*, 1, 0, 19, 0, 99*, 80, 65, 15, 65*, 63*, 21, 200, 112*, 7, 29, 38, 14, 131*, 170, 61*, 0, 67*, 66, 58, 37, 26, 1, 0, 160, 8, 18, 67, 60*, 12, 33, 0, 108, 116, 4, 75, 14, 22, 6, 2, 23, 96, 7, 2*, 96, 17, 85, 6, 34, 12, 27, 80, 33, 157, 1, 49*, 0, 28, 112, 16*, 33, 15*, 59, 7, 122*, 30*, 96, 8, 14, 0, 100, 9, 199, 11, 72*, 4, 19, 19, 14, 151*, 1, 24, 28, 5, 150, 5, 32, 32, 57, 17, 10, 151*, 15, 3, 18*, 41, 26, 121*, 20, 103, 38, 15, 110, 24, 47, 47, 105, 45, 13, 1*, 157*, 3, 0, 8, 67, 8, 13, 90, 30, 32, 0, 14, 7, 42, 7, 12, 34, 53, 77, 14, 102, 6, 25, 115, 41, 45*, 100*, 156*, 78, 61, 0, 56*, 30, 42, 19, 40, 80

England

Ian Bell

70, 65*, 162*, 6, 8, 6, 21, 59, 65, 3, 3, 0, 0, 71, 31, 115, 0, 4, 92, 9, 1, 38, 57, 18, 8, 100*, 28, 106*, 119, 4, 9, 9*, 50, 0, 60, 26, 0, 87, 7, 2, 71, 28, 109*, 3, 5, 97, 2, 11, 20, 9, 31, 0, 63, 67, 83, 74, 15, 54, 1, 34, 25, 54*, 11, 41, 9, 110, 16, 8, 21*, 0, 199, 31, 4, 50, 20, 24, 4, 17, 7, 1, 24*, 28, 4, 53, 8, 3, 72, 4, 5, 2, 140, 48, 78, 35, 5, 84, 39*, 138, 17, 128, 76, 68*, 53, 16, 1, 115, 103*, 52, 57*, 119*, 45, 0, 31, 159, 34, 235, 52, 13, 18, 61, 63*, 22, 76*, 13, 55, 11, 3*, 58, 4, 0, 22, 5, 28*, 1, 116*, 24, 26*, 11, 17, 75, 31, 6, 30, 6, 25, 109, 109, 74, 60, 4*, 6, 113, 45, 17, 5, 32, 72*, 6, 15, 60, 27, 0, 2, 16, 56, 9, 64, 8, 25, 16, 1, 167, 23, 58, 7, 143, 11, 1, 0, 0, 1, 29, 12, 1, 1, 60, 1, 11, 53, 65*, 1, 10, 13

Kevin Pietersen

57, 64*, 71, 20, 21, 0, 45, 23, 14, 158, 5, 19, 100, 42, 34, 1, 15, 87, 64, 4, 39, 7, 158, 142, 13, 41, 6, 21, 41, 38, 135, 16, 0, 96, 16, 92, 158, 2, 70, 60*, 21, 1, 41, 29, 26, 109, 226, 9, 68, 0, 28, 37, 134, 13, 19, 41, 101, 31, 18, 1, 45*, 1, 30, 42, 6, 31, 17, 129, 34, 3, 26, 42, 115, 152, 45, 13, 4, 94, 100, 13, 4, 1, 144, 97, 1, 51, 32, 41, 72*, 10, 102, 0, 49, 69, 8, 32, 44, 40, 81, 31, 0, 6, 7, 12, 99, 32, 45, 74*, 18, 10*, 64, 9, 22, 80, 6, 23, 0, 43, 227, 0, 3, 51, 36, 3, 2, 72, 85, 202*, 1, 29, 63, 63, 175, 3, 30, 151, 42*, 32, 13, 80, 78, 42, 16, 149, 12, 17, 2, 186, 54, 0, 73, 6, 0, 12, 73, 14, 64, 2, 5, 113, 8, 26, 44, 50, 62, 18, 26, 4, 53, 19, 45, 71, 49, 3, 6

Matt Prior

126*, 21, 75, 40, 0, 62, 1, 42, 11, 7, 0, 12*, 0, 63, 79, 4, 19*, 53*, 33, 2, 64, 0, 39, 15*, 131*, 61, 42, 63, 56, 14, 8, 61, 41, 37*, 22, 18, 4, 4, 0, 60, 76, 4, 14, 0, 0*, 7, 62, 16, 93, 6, 102*, 15, 84*, 5, 22, 0, 27*, 12, 10, 85, 118, 126, 4, 0, 71, 103*, 1, 73, 5, 18*, 7, 41, 11, 19, 16, 60, 40, 68, 7, 27, 73, 48, 91, 21, 41, 57, 23, 23*, 82, 73, 110*, 0, 0, 39, 4*, 1, 31, 6, 1*, 30, 17, 0, 47, 0*, 0, 4, 0, 69, 8, 26, 86, 16, 27*, 10, 5, 23, 12

Andrew Strauss

112, 83, 62, 10, 0, 6, 137, 35, 24, 5, 90, 12, 14, 0*, 126, 94*, 25, 136, 45, 39, 147, 0, 44, 0, 69, 8, 2, 37, 48, 6, 6, 106, 35, 23, 129, 1, 9, 23, 12, 0, 28, 46, 18, 13, 128, 4, 48, 30, 16, 7, 55, 30, 128, 42, 36, 116, 38, 54, 12, 11, 14, 34, 42, 0, 50, 31, 29, 24, 33, 24, 15, 6, 0, 77, 13, 96, 18, 4, 55, 6, 32, 43, 2, 8, 44, 0, 177, 63, 60, 106, 37, 44, 27, 0, 20, 25, 6, 58, 123, 108, 0, 21*, 7, 9, 6*, 169, 14, 142, 38, 142, 14, 16, 14*, 26, 30, 17, 161, 32, 69, 3, 32, 55, 75, 46, 1, 54, 2, 45, 0, 22, 83, 82, 21, 45, 0, 25, 53*, 15, 4, 13, 0, 110, 1, 52, 15, 69, 60, 20, 4, 0, 3, 22, 32, 32, 16, 87, 40, 26, 27, 61, 0, 122, 1, 141, 45, 17, 0, 27, 37, 22, 20, 1

Graham Thorpe

6, 114*, 0, 13, 37, 60, 16, 14, 0, 20, 86, 3, 7, 84, 9, 72, 73, 79, 15*, 28, 67, 51, 9, 10, 47*, 26, 83, 123, 0, 20, 61, 52, 42, 30, 0, 94, 0, 19, 76, 74, 38, 13, 34, 17, 2, 27, 12*, 20, 59, 21, 17*, 89, 21, 45, 77, 3, 16, 54, 9, 13, 2, 5, 50*, 119, 108, 18, 2, 138, 21, 30*, 3, 7, 15, 15, 53, 82*, 27, 62, 0*, 8, 39, 32, 19, 10, 3, 103, 36*, 5, 84*, 10, 43, 10, 0, 0, 0, 77, 9, 6, 21*, 7, 7, 27, 25*, 10, 44, 0, 46, 40, 10, 118, 5, 79, 0, 18, 64*, 7, 12, 59, 46, 113*, 32*, 80, 138, 10, 20, 2, 23, 62, 17, 200*, 11, 1*, 42, 3, 27, 65, 123, 32, 4, 1, 124, 64, 18*, 0, 54, 43, 10, 57, 41, 13, 19, 19, 90, 13*, 119*, 10, 23*, 3, 51*, 34, 45, 104*, 19, 38, 61, 54, 114, 4, 31*, 1, 118*, 12, 26, 0, 1, 86, 8, 42*, 66*

Marcus Trescothick

66, 38*, 1, 78, 7, 71, 1, 30, 10, 13, 24, 122, 57, 23, 13, 23, 10, 36, 10, 117, 0, 76, 15, 3, 69, 32, 37, 10, 55, 24, 66, 46, 99, 12, 8, 9*, 0, 33, 37, 88, 0, 14, 13, 76, 161, 81, 23*, 57, 58*, 72, 1, 35, 0, 34, 4, 37, 37, 19, 22, 59, 43, 31, 52*, 6, 23, 24, 0, 59, 4, 219, 69*, 113, 32, 60, 1*, 23, 24, 36, 14, 70, 0, 7, 6*, 1, 4, 2, 42, 16, 88, 86, 2, 132, 30*, 63, 9, 16, 45, 105, 107, 0, 12, 30, 4*, 47, 0, 18, 132, 28, 0, 16, 180, 20, 7, 194, 151, 4, 44, 90, 21, 63, 41, 65, 27, 43, 33, 193, 5, 48, 0, 50, 0, 106, 27, 0, 24, 31, 16, 18, 5, 28, 58, 6, 4

Jonathan Trott

41, 119, 28, 69, 18, 20, 42, 5, 8, 39, 14, 64, 19, 226, 36*, 3, 38, 26, 55, 53*, 12, 36, 184, 29, 135*, 78, 4, 31, 168*, 0, 203, 2, 58, 4, 70, 22, 4, 2, 12, 112, 64, 5, 58, 13, 35, 17*, 17, 71, 10, 35, 30*, 8, 63, 0, 17, 0, 87, 3, 44, 143, 45, 52, 121, 27, 37, 39, 56, 28, 76, 48, 0, 58, 0, 5, 11, 49, 23, 40, 59, 10, 9, 0, 4, 59, 0, 0, 9

Michael Vaughan

33, 5, 21, 29, 42, 5, 69, 4, 41, 29, 76, 10, 9, 26, 8, 32, 120, 14, 11, 31*, 64, 27, 0, 7, 34, 27, 36, 64, 115, 46, 36, 24*, 0, 100, 197, 61, 15, 195, 47*, 33, 0, 177, 41, 34, 9, 11, 145, 0, 183, 8, 20, 156, 22, 33, 29, 1, 5, 15, 21, 23, 13, 48, 81*, 54, 25, 24, 8, 52, 105, 18, 14, 15, 11*, 0, 23, 17, 32, 7, 140, 13, 61, 10, 103, 101*, 12, 3, 12, 33, 66, 10, 15, 18, 10, 11, 20, 82*, 54, 0, 26*, 120, 44, 3, 4, 24, 1, 166, 14, 58, 0, 11, 45, 2, 9, 58, 13, 103, 41, 40, 19, 48*, 79, 30, 9, 124, 11, 42, 37, 5, 87, 61, 1, 24, 63, 9, 32, 13, 2, 4, 106, 30, 48, 16, 2, 0, 21, 0, 17

India

Rahul Dravid

95, 84, 8, 40, 24, 34, 31, 23, 7, 56, 7, 27*, 2, 12, 148, 81, 43, 51*, 57, 78, 2, 37*, 92, 69, 2, 6, 34, 92, 93, 85, 52, 56, 86, 23, 6, 118, 44, 0, 28, 190, 103*, 53, 10, 33, 29, 24, 13, 107, 1, 144, 48, 1, 33, 12, 35, 6, 9, 14, 29, 0, 22, 37, 17, 18, 28, 41*, 200*, 70*, 162, 9, 39, 25, 180, 81, 4, 44, 68*, 26, 12, 61*, 15, 75, 36, 36, 2, 11, 2, 87, 86, 7, 26*, 3, 65, 1, 6, 144*, 67, 36, 17, 14, 91, 5, 30, 46, 63, 13, 115, 148, 217, 100*, 11, 6*, 14, 17, 76, 7, 9, 39, 222, 73, 13, 5, 1, 43*, 233, 72*, 49, 92, 38, 91*, 6, 33, 0, 270, 0, 60, 26, 21, 2, 31*, 27, 54, 80, 47*, 0, 160, 50, 110, 135, 22, 16, 77, 98, 0, 23, 32, 24, 53, 128*, 103, 5*, 3, 2, 40, 71, 95, 42*, 52, 9, 49, 62, 146, 22, 68*, 81, 68, 32, 1, 11, 5, 29, 47, 61, 2, 129, 2, 9, 37, 11*, 55, 12, 38, 34, 50, 8*, 19, 42, 5, 16, 53, 38, 93, 3, 18, 11*, 111, 3, 17, 29, 18*, 14, 10, 2, 44, 10, 68, 51, 5, 39, 11, 11, 0, 3, 3, 4, 136, 0, 66, 8*, 83, 62, 35, 60, 177, 38, 144, 74, 4, 24, 111*, 111*, 18, 44, 3, 23, 7, 77, 13, 1, 21*, 104,

1, 45, 191, 14, 43, 25, 2, 5, 31, 40, 112, 5, 55, 5, 34*, 103*, 36, 117, 6, 22, 18, 146*, 13, 54, 31, 119, 82, 33, 68, 10, 5, 29, 9, 47, 1, 25

Sourav Ganguly

131, 136, 48, 66, 21*, 6, 0, 39, 41, 16, 0, 23, 30, 73, 60, 42, 6, 22, 8, 0, 147, 45, 109, 99, 173, 11, 3, 30*, 65, 17, 16, 47, 36, 5, 48, 11, 101*, 54, 2, 13, 62*, 17, 24, 56, 78, 2, 64*, 0, 125, 53, 60, 43, 31, 17, 1, 25, 2, 31, 1, 13, 84, 27, 65*, 30, 8, 1, 23, 48, 22, 4, 5, 9, 0, 15, 4, 18, 98*, 1, 30, 14, 30, 42, 4*, 47, 5, 16*, 0, 38, 136, 20, 5, 25, 75*, 48, 60*, 45, 36, 28, 5, 0, 68, 99, 128, 51, 4, 0, 29, 16, 17, 2, 5, 5, 100*, 25, 144, 2, 12, 37, 73, 16, 77, 45, 5, 9, 57, 40, 71, 88, 21, 12, 12, 1, 2, 101, 16, 5, 40, 39, 34, 37, 51*, 25, 0, 26, 66, 46, 100, 13, 15, 34, 40, 79, 2*, 37, 57, 8, 48, 102, 46, 239, 91, 43, 40, 67, 51, 9, 0, 7, 18, 24, 0, 87, 87, 13*, 23, 4, 0, 16, 35, 18, 47, 26*, 102, 27, 5, 32*, 85, 0

VVS Laxman

11, 51, 14, 1, 5, 35*, 0*, 64, 27, 0, 6, 19, 56, 95, 6, 15, 23, 0, 35, 8, 5, 67, 11, 25, 41, 0, 5, 1, 7, 167, 16, 0, 18*, 20, 12, 59, 281, 65, 66, 28, 38, 15, 20, 32, 29, 89, 28, 75, 12, 13, 69, 69*, 74, 1, 43, 130, 65*, 23, 43*, 74, 22, 14, 6, 40, 45, 24, 48, 154*, 0, 0, 23, 4, 64, 44, 104*, 67*, 75, 24*, 148, 32, 19, 18, 178, 29, 11, 13, 71, 31, 3, 4, 13, 2, 1, 69, 9, 38, 32, 9, 58, 0, 24, 79*, 5, 140, 8, 5, 69, 11, 104, 5, 0*, 90, 8*, 19, 21, 0, 0*, 29, 31, 0, 100, 63, 18, 16, 28, 73, 50*, 15, 13, 1, 15, 39, 54, 51, 46*, 72*, 6*, 112*, 5, 14*, 26, 42, 109, 20, 27, 79, 51, 12, 39, 3, 35, 50, 56, 21, 39, 13, 25, 61*, 0, 42*, 12, 200*, 59*, 64, 4, 24, 26, 0, 15, 30, 76, 124*, 4, 61, 0, 51*, 63, 62, 7, 69*, 143*, 22, 69, 29, 56, 103*, 2, 73*, 40, 91, 74, 12, 7, 8, 38, 96, 15, 32*, 12, 0, 85, 87, 56, 3*, 10, 56, 54, 4, 30, 2, 2, 24, 1, 58*, 176*, 32, 31, 2, 1, 2, 66, 31, 0, 18, 35

Virender Sehwag

105, 31, 13, 20, 66, 74, 84, 27, 106, 0, 8, 12, 147, 61, 33, 35, 10, 2, 12, 1, 25, 29, 17, 130, 1, 45, 0, 47, 47, 195, 11, 72, 47, 309, 39, 90, 0, 39, 0, 155, 12*, 22, 58, 8, 5, 164, 88, 10, 13, 10, 173, 36, 81, 15, 201, 38, 44, 44, 14*, 76, 7, 36, 20, 0, 254, 31, 5, 4, 2, 0,

11, 76*, 6, 0, 36, 41, 180, 31, 65, 0, 4, 4, 33, 0, 8, 40, 4, 29, 43, 63, 151, 319, 6, 17, 8, 22, 25, 13, 201*, 50, 21, 34, 45, 6, 35, 90, 1, 16, 66, 92, 9, 83, 0, 17, 24, 34, 22, 48, 12, 16, 51, 131, 293, 52, 45, 56, 0*, 109, 16, 165, 109, 31, 99, 109, 0, 59, 17, 30, 7, 173, 1, 96, 54*, 74, 0, 63, 25, 32, 13, 11, 0, 0, 8, 33, 55, 55, 38, 37, 60, 67, 7, 30, 4, 0, 10, 18, 62, 47, 43, 38, 117, 25, 30, 9, 23, 49, 0, 2, 19, 6

Sachin Tendulkar

15, 59, 8, 41, 35, 57, 0, 24, 88, 5, 10, 27, 68, 119*, 21, 11, 16, 7, 15, 40, 148*, 6, 17, 114, 5, 0, 11, 111, 1, 6, 0, 73, 50, 9*, 165, 78, 62, 28, 104*, 71, 142, 96, 6, 43, 11*, 34, 85, 179, 54, 40, 10, 4, 0*, 52*, 2, 24, 122, 31, 177, 74, 10, 0, 42, 7, 18, 2, 61, 36, 15, 4, 169, 9, 35, 9, 7, 15*, 88, 92, 4, 83, 143, 139, 8, 23, 15, 148, 13, 4, 155*, 79, 177, 31, 34, 7, 47, 113, 67, 0, 136, 6, 29, 0, 9, 53, 124*, 18, 126*, 15, 44*, 217, 15, 61, 0, 116, 52, 45, 4, 97, 8, 21, 20, 18, 122, 39, 201*, 76, 65, 10, 10, 126, 17, 74, 36*, 20, 69, 155, 15, 1, 22*, 88, 103, 26, 90, 176, 36, 42, 79, 117, 0, 0, 8, 0, 41, 86, 16, 12, 34, 92, 193, 54, 35, 43, 16*, 36, 176, 8, 51, 9, 32, 8, 7, 55, 1, 0, 1, 37, 0, 44, 241*, 60*, 194*, 2, 8, 1, 8, 2, 5, 55, 3, 20, 32*, 248*, 36, 94, 52, 52, 41, 16, 22, 109, 16, 23, 19, 14, 23, 26, 16, 28*, 4, 1, 34, 44, 14, 63, 0, 64, 14, 101, 31, 122*, 37, 16, 91, 1, 82, 1, 1, 56*, 82, 62, 15, 154*, 12, 71, 13, 153, 13, 0, 27, 12 5, 31, 6, 14, 13, 49, 88, 10*, 68, 47, 109, 12, 37, 103*, 11, 5, 160, 49, 64, 62, 9, 4, 100*, 40, 53, 105*, 16, 143, 7, 100, 106, 8, 84, 203, 41, 54, 98, 38, 214, 53*, 40, 12, 13, 61, 36, 111*, 13, 6, 146, 14*, 34, 12, 16, 56, 1, 40, 23, 91, 7, 76, 38, 94, 3, 73, 32, 41, 80, 15, 8, 25, 13, 19, 17, 27, 13, 8, 8, 76, 5, 2, 81, 13*, 7, 37, 21, 32, 1, 10, 74

New Zealand

Stephen Fleming

16, 92, 54, 11, 41, 39, 14, 11, 48, 15, 4, 31, 79, 53, 56, 47, 30, 17, 27, 35, 0, 66, 16, 41, 25, 0, 49, 21, 84, 3, 1, 22, 39, 56*, 19, 92*, 67, 4, 129, 9, 1, 0, 62, 11, 51, 2, 59, 52, 27, 27, 75, 91, 0, 10, 4, 0, 0, 36, 19, 78, 174*, 14, 10, 78, 3, 42, 17, 0, 18, 27, 25, 1,

5*, 38, 66*, 4, 43, 73, 2, 31, 48, 64*, 66, 67, 21, 8, 16, 60, 30, 2, 11, 12, 9, 57, 99, 14, 8, 14, 22, 55, 86, 5, 32, 51*, 0, 57, 71, 105, 4, 4, 61, 12, 48, 3, 11, 1, 1, 2, 66, 130, 34, 6, 5, 25, 21, 32, 274*, 69*, 0, 33, 1, 8, 30, 192, 0, 0, 24, 27, 4, 31*, 30, 9, 34, 4, 97, 11, 117, 45, 29, 202, 0, 11, 83, 3, 18, 17, 0, 1, 65, 3, 16, 41, 88, 73, 65, 14, 33, 97, 0, 6, 262, 46, 37, 48, 0, 0, 27, 40, 17, 43, 54, 14, 87, 41, 66, 34, 31, 59, 66

Mark Richardson

6, 13, 99, 23, 77, 26, 60, 46, 75, 1, 59, 46, 73*, 106, 26, 57, 30, 9, 30, 143, 83, 2, 76, 60, 4, 5, 25, 8, 32, 41, 0, 95, 71, 89, 14*, 13, 28, 85, 6*, 55, 55, 6, 21, 145, 44, 15, 82, 41, 4, 45, 10, 14, 37, 93, 101, 13, 40, 73, 49, 15, 28, 19, 4, 9, 16

Jesse Ryder

1, 38, 91, 39*, 30, 24, 13, 3, 89, 57, 59*, 102, 21, 201, 3, 0, 42, 24, 23, 38, 103, 70, 20, 59, 22, 22, 0, 0, 17, 6, 36, 0, 16

Pakistan

Inzamam ul-Haq

8*, 0, 8, 26, 5, 19, 23, 75, 10, 6, 7, 26, 123, 21, 57*, 38, 14, 33, 43, 20*, 135*, 5, 20, 81, 7*, 100*, 9, 58*, 14, 0, 66, 3, 19, 95, 71, 65, 47, 101, 83, 95, 50, 26, 21, 0, 5, 62, 27, 40, 39, 59, 0, 82, 148, 70, 2, 65, 35, 0, 14, 1, 12, 43, 54*, 8, 56, 96, 5, 92*, 177, 4, 0, 6, 4, 24, 12, 13, 10, 14, 0, 97, 9, 21*, 19, 2, 10, 51, 26, 6, 0, 4, 88, 12, 12, 118, 22, 8, 44, 20, 58*, 9, 86, 138, 135, 8, 29, 55, 68, 12, 13, 112, 63, 0, 71, 142, 27, 130, 5, 20, 13, 20, 114, 85, 105*, 43, 30, 29, 99, 329, 39, 112, 11, 18, 13, 32, 60, 0, 35*, 43, 10, 138*, 23, 60, 51, 34, 72*, 77, 0, 118, 15, 9, 32, 3, 117, 1, 0, 57, 86, 30, 13, 184, 31*, 50, 117*, 1, 0, 53, 72, 109, 100*, 97, 1, 119, 31, 48, 15, 69, 56*, 0, 13, 26, 37, 31, 0, 31, 10, 18, 58*, 42, 35, 92*, 1, 6, 22, 14, 3

Mohammad Yousuf

5, 1, 60, 64, 9, 52, 28, 9, 11, 75, 14, 120*, 53, 26, 3, 0, 2, 56, 83, 0, 95, 75, 17, 2, 18, 0, 32, 18, 8, 88, 7, 11, 0, 115, 19, 103*, 42, 2, 11, 41, 124, 77, 117, 24, 51, 42, 203, 0, 16, 26, 6, 4, 49, 102*, 72, 204*, 6, 7, 29, 63, 0, 159, 12, 42, 0, 50, 46, 15*, 64*, 8, 28, 60, 88*, 35, 112, 72, 13, 48, 17, 44, 46, 1, 1, 27, 111, 12, 8, 30, 6, 68, 104, 22, 37, 5, 16, 78, 20, 223, 173, 65, 126, 0, 97, 17, 14*, 202, 48, 38, 15, 192, 8, 128, 192, 56, 191, 102, 124, 32, 18, 83, 18, 25, 63*, 27, 18, 6, 44*, 24, 10*, 112, 12, 10, 6, 90, 23, 17, 41, 0, 83, 0, 89, 22, 61, 46, 19, 7, 23, 56, 33, 0, 10

South Africa

Herschelle Gibbs

31, 9, 17, 5, 0, 25, 31, 7, 54, 1, 37, 2, 4, 2, 4, 35, 49, 42, 25, 2, 51, 34, 211*, 120, 0, 85, 48, 10, 2, 26, 29, 3, 47, 46, 4, 0, 1, 8, 83*, 34, 87, 34, 19, 85, 45, 18, 51, 147, 74, 107, 1, 196, 12, 78, 9, 14, 21, 32, 10, 34, 47, 12, 39, 51, 104, 41, 114, 92, 7, 11, 25*, 228, 17, 21, 179, 9, 49, 19, 28, 0, 2, 183, 9, 27, 59, 98, 20, 60, 6*, 142, 33, 142, 192, 8*, 40, 47, 80, 61, 77, 16, 0, 4, 15, 36, 4, 24, 161, 98, 14, 4, 8*, 47, 5, 49, 34, 8, 23, 21, 33, 94, 9, 27, 67, 18, 0, 9, 17, 16, 53, 6, 2, 19, 18, 0, 92, 0, 0, 63, 9, 7, 0*, 94, 2, 40, 54, 18, 13, 16, 63, 8, 25, 0, 0, 27

Jacques Kallis

1, 7, 6, 39, 0, 2, 2, 61, 15, 101, 16, 45, 15, 15, 15, 43, 22, 10, 69, 3, 49, 12, 0*, 61, 0, 132, 47, 11, 40, 3, 53, 57*, 30, 3, 11, 23*, 110, 88*, 83, 27, 7, 148*, 17, 4, 64, 115, 12, 1, 85*, 0, 69, 105, 25, 5, 36*, 95, 29, 40, 16, 87, 19, 0, 160, 13, 12, 23, 79*, 21, 15, 49, 7, 50, 30, 53, 0, 11, 20, 5, 30*, 17, 51, 157*, 42*, 189*, 68, 21*, 24, 89*, 5, 65*, 38, 99, 4, 34, 3, 8, 23, 73, 16, 61*, 75*, 139*, 75, 84, 6, 105, 31, 27, 13, 6, 41, 66, 35, 29, 18, 10, 43, 158, 44, 177, 73, 130*, 130*, 92, 150*, 40, 71, 0, 1, 59, 52*, 13, 3, 37, 28*, 121, 55, 0, 61, 162, 10, 149, 66, 33, 0, 8, 136*, 54, 58, 0, 109*, 39, 19*, 78, 147, 44, 39*, 23, 9, 111, 50*, 6, 36, 114, 7, 37, 27, 38, 62, 71, 9, 13, 12, 27, 54, 32, 18, 60*,

24, 91, 28, 51, 155, 100*, 59, 107*, 29, 186, 131, 0, 85, 36, 22*, 74, 17, 7, 39*, 13, 19, 132, 1, 15, 7, 13, 4, 64, 5, 2, 9, 16, 24, 63, 57, 26, 37, 4, 27, 45, 22, 93, 102, 120, 4, 75, 3, 108, 46, 7, 173, 10, 20, 28, 40, 110, 62*, 43, 0*, 201*, 10, 17, 161, 109*, 0, 2*, 54, 2, 31, 0, 0, 224, 0, 113, 6, 182*, 19, 27, 3, 31, 147, 49, 58, 46, 2, 37, 60, 8, 50, 7, 2, 21, 0, 34, 115

Gary Kirsten

16, 67, 41, 43, 7, 47, 35, 29, 10, 41, 72, 44, 7, 65, 2, 0, 9, 33, 29, 66*, 64, 25, 62, 42, 16, 76, 1, 13, 110, 1, 8, 51, 69, 23, 41*, 17, 20, 102, 133, 43, 7, 2, 2, 103, 0, 29, 1, 9, 8, 0, 43, 16, 6, 98, 56, 100*, 4, 83, 0, 11, 0, 77, 108*, 3, 20*, 0, 25, 38, 44, 62, 15, 13, 75*, 12, 4, 9*, 210, 7, 6, 6, 3, 62, 7, 29, 2, 26, 71*, 0, 5, 0, 134, 128, 65, 40, 12*, 13, 15, 2, 11, 275, 80, 0, 50, 20, 79, 12, 55, 0, 13, 11, 40, 31, 1, 49, 47*, 10, 180, 34, 52, 150, 24, 23, 22, 0, 0, 8, 9, 0, 14, 220, 31*, 65, 73, 30*, 4, 5, 47, 7, 10, 10, 18, 153, 1, 12, 7, 87, 21, 64, 150, 160, 55, 11, 11, 56, 19, 44, 1, 108, 130, 60, 90, 29, 53*, 46, 54, 118, 137, 16, 10*, 10, 137, 34*, 1, 1, 1, 76

Ashwell Prince

49, 28, 10, 20, 0, 48, 2, 0, 3, 20, 5, 139*, 45, 23, 131, 28, 8, 6, 26, 119, 18, 17, 27, 33, 7, 93, 9, 9, 11, 108*, 4, 43*, 1, 61, 86, 17, 24, 97, 121, 0, 26, 38*, 138, 2, 22, 19, 59*, 36, 45, 63, 11, 1, 25*, 13, 20, 10, 98, 12*, 123*, 10, 38, 2, 23, 5, 2, 16, 22*, 101, 9*, 149, 39, 2, 4, 24, 59*, 162*, 150, 45, 0, 2, 16, 0, 15, 19, 0, 1, 23, 57, 16*, 9, 78*, 13, 39*, 47, 22, 0, 50, 2, 39, 11, 7

Graeme Smith

3, 68, 1, 42, 200, 24, 73, 15, 0, 16, 13*, 151, 16, 15, 277, 85, 259, 35, 5, 2, 14, 18, 19, 33, 12, 2, 65, 132, 44, 14, 42, 24, 139, 23*, 25, 5, 88, 0, 47, 125*, 23, 74, 65, 17, 37, 47, 0, 71, 0, 55, 9, 5, 74, 2, 29, 67*, 25, 3, 121, 41, 2, 34, 148, 41, 104, 126, 50*, 12, 0, 34, 30, 22, 25, 39, 5, 19, 16, 0, 40, 45, 7, 25, 63, 68, 5, 10, 5, 58, 94, 55, 0, 32, 28, 10, 64, 33, 42, 25, 46, 133, 1, 9, 2, 28, 11, 28, 85, 147, 10, 62, 232, 73, 35, 34, 69, 35,

8, 107, 44, 3*, 7, 154*, 46, 0, 157, 27, 48, 108, 62, 75, 30*, 3, 0, 69, 2*, 0, 12, 75, 22, 30, 183, 105, 6, 4, 20, 23, 90, 132, 46, 70, 10, 62, 9, 37, 6, 29, 37, 101*, 11, 36, 61, 15, 26, 16, 0*, 53, 115, 13, 55*, 5, 41, 131, 52, 52, 14, 23, 10, 23, 122, 0, 16, 84, 1, 54, 24, 52, 19, 29, 5, 68, 44, 47, 27*, 10, 4, 9, 14, 5, 3

Sri Lanka

Tillakaratne Dilshan

9, 163*, 0, 37, 13, 7, 31, 5, 6, 28*, 5, 17, 36, 0, 5, 10, 63, 100, 83, 104, 6, 0, 43, 0, 31, 10, 17*, 14, 35, 21, 25, 1, 3, 23*, 28, 9, 73, 32, 27*, 36, 49, 86, 168, 8*, 0, 32, 65, 65, 22, 33, 69, 8*, 22, 11, 0, 69, 27, 59, 8, 32, 45, 4, 18, 79, 0, 17*, 84, 20, 4, 62, 25, 125*, 0, 38, 23, 14, 47, 162, 143, 0, 8, 145, 28, 22, 20, 44, 92, 123*, 29, 33, 112, 0, 11, 109, 16, 25, 68*, 54, 14, 41, 13, 0, 54, 4, 26, 50, 10, 193, 4, 12, 4, 36, 83, 6, 6, 47, 4, 78, 5, 11, 0, 14, 35, 101, 56, 121, 28, 5, 14, 147, 11, 11, 0, 34, 5, 54, 126, 0, 57

Sanath Jayasuriya

35, 18, 12*, 11, 66, 77, 35*, 81, 45, 19, 1*, 2, 4, 6*, 0, 31*, 44, 16, 65, 0, 22, 1, 9, 1, 10, 48, 112, 0, 41, 18*, 0, 50, 20, 3, 31, 62, 72, 113, 85, 0, 90, 17, 340, 32, 199, 53, 17, 50, 37, 6, 0, 5, 68, 17, 0, 51, 16, 10, 59, 21, 13, 8, 213, 24*, 18, 18, 0, 21*, 0, 49, 6, 7, 4, 16*, 17, 56, 30, 6, 24, 10, 26, 8, 32, 21, 188, 148, 28, 0, 85, 17, 0, 26, 8, 0, 16, 16, 14, 9, 0, 45, 23, 111, 6*, 3, 6, 30, 89, 25, 6*, 16, 55, 85, 8*, 92, 139, 28, 36, 88, 1, 18, 8, 12, 35, 26, 145, 85, 32, 0, 50, 82, 9, 8, 72*, 26, 13, 48, 17, 32, 27, 85, 35, 5, 1, 131, 71, 51, 157, 48, 8, 16, 13, 22, 12, 74, 43, 19, 38, 253, 26, 107, 48, 5*, 22, 2, 3, 15, 2, 36, 46, 13, 6, 13, 14, 4, 4, 4, 47, 73, 5, 10, 0, 31, 7, 39, 3, 45, 10, 78

Mahela Jayawardene

66, 16, 7, 52, 54, 167, 16, 11, 9, 242, 4, 50, 46, 9, 46, 21, 17, 91, 6*, 2, 42, 35, 36, 10, 1, 29, 77, 1, 72, 9, 2, 167, 18, 1, 34, 101*, 98, 7, 0, 45, 17, 23, 61, 101, 18, 71, 11, 28, 104, 25, 139, 150, 99, 88, 16, 39, 18, 56, 76, 17*, 68, 12*, 107, 14*, 47, 59, 17, 28, 0,

39, 1, 44, 40, 58, 15, 32*, 45, 10, 32, 17, 86*, 45, 52, 134, 68, 21, 17, 13, 29, 37, 37, 100*, 14, 44, 43, 6, 237, 5, 82, 3, 0, 57, 16, 32, 141, 1, 13, 3, 41*, 6, 43, 63, 2, 71, 60, 67, 0, 57, 30, 23*, 49, 1, 82, 4, 15, 61, 119, 0, 5, 0, 45, 374, 13, 123, 8, 0, 0, 31, 127, 49, 165, 14, 49, 104, 0, 1, 65, 195, 213*, 136, 33, 26, 12, 136, 86, 5, 2, 50*, 3, 166, 11, 22, 240, 22, 30, 30, 0, 19, 37*, 79, 2, 114, 27, 92, 96, 275, 47, 10, 29, 12, 48, 174, 5, 56, 5, 59, 58, 2, 4, 15, 49, 25, 4, 6, 11, 105, 4, 51, 51, 30, 15, 31, 14, 30, 12, 180, 5, 105, 64, 62, 14, 0, 1*, 12, 11, 91, 4, 5, 12, 19, 3, 0, 72, 60, 203*, 72, 11, 55, 18, 22, 79, 3, 10, 165, 0, 59, 26, 4, 54

Hashan Tillakaratne

0, 6, 0, 55, 21, 12, 26, 31, 3, 20, 16, 49, 42*, 11, 14, 82, 1, 93, 93*, 36*, 28, 2, 51, 86, 92, 33*, 9, 9, 37, 0, 9*, 7, 47, 0, 80, 5, 40, 34, 8, 9, 83*, 1, 1, 15*, 116, 9, 74, 36, 108, 44*, 48, 115, 0, 24, 50, 6, 119, 14, 38, 65, 3, 20, 126*, 55*, 8, 2, 10, 103, 54, 10, 24*, 1*, 14, 9, 25, 18*, 44, 7, 0, 22, 13, 55, 0, 10, 43, 40, 0, 14, 9, 11, 10, 16, 136*, 10*, 105*, 87, 7*, 204*, 96, 37, 3, 19*, 17*, 20, 39, 20, 32*, 18, 5*, 24, 27, 104*, 6, 144, 93, 13, 13, 7, 0, 1, 45, 20, 12, 33, 25, 16, 7, 74*, 17

Thilan Samaraweera

103*, 77, 29, 3*, 87, 123*, 17, 76, 8, 58, 11, 45, 1, 3, 23*, 142, 36*, 15, 41, 53, 6, 32*, 1, 32, 70, 0, 13, 19, 21, 21*, 100, 21, 13, 22, 88, 73, 17, 11, 51, 37, 0, 78, 138, 35*, 1, 0, 1, 5, 58, 20, 4, 64, 65, 4, 0, 6, 3, 8, 13, 20, 0, 56*, 6, 125, 127, 14, 67*, 35, 90, 62, 19, 77, 231, 24*, 214, 31, 34, 21, 6*, 6, 73, 159, 20, 143, 25, 70, 2, 78*, 1, 0, 0, 76*, 10*, 137*, 83, 52, 19*, 80, 58, 0, 9, 17*, 31, 87*, 26, 0, 17, 43, 36, 32, 102, 43, 11, 115*, 20, 36, 54, 47, 6, 15, 0, 73, 10, 17, 76, 7, 7, 49, 10, 1, 12, 0

Kumar Sangakkara

23, 24, 5, 25, 6, 74, 17, 32, 11, 3, 98, 58, 17, 95, 45, 0, 105*, 31, 13, 47, 54, 140, 15, 45, 55, 128, 42, 29, 56, 230, 14*, 10, 6*, 16, 1, 40, 32, 75, 26, 7, 35, 89, 67, 10, 27*, 56, 75, 12, 71, 19, 34, 10, 31, 22, 7, 5, 29, 22, 27, 11, 270, 2, 0, 74, 66, 58, 13, 232,

64, 2, 59, 13, 138, 5, 16, 45, 34, 0, 6, 157*, 30, 5, 30, 3, 33, 41, 17, 69, 46, 0, 8, 185, 79, 16, 21, 65, 25, 18, 36, 66, 287, 14, 39, 4, 100*, 156*, 8, 6, 200*, 222*, 57, 192, 92, 152, 1, 46, 50, 21, 10, 14, 12, 68, 1, 144, 4, 43, 67, 5, 54, 70, 65, 104, 9, 14, 87, 46, 45, 130*, 8, 46, 50, 109, 31, 44, 11, 18, 137, 103, 219, 42*, 75, 28, 73, 4, 150, 1*, 11, 14, 26, 12, 2, 119, 10, 17, 48, 69, 79, 1, 2, 0, 108, 35, 34, 0, 14, 0, 21, 199*, 1, 192, 24*, 0, 74*, 5, 0, 16, 4, 63, 58, 27*, 142, 105, 139, 55, 75, 319, 105, 147, 61, 79, 55, 24, 76, 0, 72, 221, 21, 22, 59, 6, 1, 203, 5, 50, 18, 34, 0, 5, 40, 32, 18

West Indies

Shivnarine Chanderpaul

62, 19, 50, 77, 5, 75*, 4, 11*, 69, 61*, 18, 5*, 80, 82, 41, 8, 82, 14, 48, 71, 58, 40, 20, 8, 3, 52, 48, 42, 79, 137*, 3, 24, 58*, 0, 14, 95, 7, 21, 16, 34, 0, 28, 39, 118, 0, 45, 3*, 5, 74, 1, 4, 16, 4, 75, 6, 5, 38, 43, 14, 0, 5, 70, 12, 49, 12, 46*, 9, 16, 89, 31, 73, 22, 9, 18, 62*, 40, 16, 7, 7, 7, 74, 140, 1, 67*, 101*, 136*, 58, 59, 35*, 17, 51, 54, 36*, 27, 3, 140, 4, 16, 19*, 100, 31, 0, 21, 1, 104, 36, 39, 15, 15, 34, 74, 0, 109, 42, 27, 7, 0, 2, 42, 50, 0, 7, 101*, 128*, 97*, 45, 43, 76, 2, 14, 32, 203*, 35, 1, 53, 31, 127, 92, 153*, 28, 0, 69, 48*, 13, 24, 2, 7, 39, 10, 25, 4, 13, 15, 8, 36, 2, 24, 62, 30, 54, 97*, 11, 10, 13, 5, 81, 14, 36, 69, 74, 50, 116*, 136*, 70, 104, 8, 65*, 70*, 0, 23, 3, 18, 86*, 118, 11, 107*, 77*, 79*, 50, 76, 126*, 0, 20, 1, 55, 70, 147*, 6, 0, 4, 23, 47, 2, 2, 62, 27, 26, 15, 166, 22, 71*, 32, 8, 0*, 54, 27, 36*, 23, 30, 37, 12, 23, 116*, 49, 18, 59*, 118, 47, 4, 47, 103*, 12, 94, 68, 69, 87*, 91, 46, 11, 0, 9, 43*, 203*, 1, 150*, 26, 108, 36, 31*, 25, 41, 76, 1, 6, 31*, 122*, 20, 84*, 24, 47, 15, 25, 85*, 84*, 101*, 21, 4, 7, 9, 50, 46, 13, 1, 7, 25, 0

Chris Gayle

33, 0, 13, 13, 0, 81, 44, 10, 23, 40, 48, 11, 12, 25, 32, 175, 6, 52*, 9, 1, 44, 0, 0, 0, 12, 13, 52, 14, 0*, 32, 68, 15, 3, 73, 204, 7, 42, 23, 0, 88, 51, 38, 37, 71, 56, 0, 19, 27, 31, 0, 14, 13, 47, 0, 8, 26, 116, 32, 77, 107, 5, 9, 62, 16, 6, 15, 69, 141, 66*, 14, 66, 81, 7,

82, 5, 42, 12, 105, 6, 1, 0, 5, 317, 4, 50, 33, 15, 10, 33, 56, 4, 25, 82, 30, 68, 30, 72, 69, 46, 2, 83, 3, 0, 0, 34, 11, 93, 40, 2, 30, 47*, 11, 13, 23, 16, 28, 52, 66, 29, 46, 38, 0, 51*, 45, 10, 14, 26, 74, 34, 197, 104, 30, 46, 6, 102, 4, 28, 0, 19, 54, 31, 1, 26, 165*, 102, 21, 6, 73, 50, 20, 10, 333, 30, 3, 0, 150, 64*, 8, 8, 24, 19, 25, 20*, 40, 4*, 101, 18, 33, 11, 35, 64, 10, 1, 80*, 42, 11, 64, 9*

Brian Lara

44, 5, 17, 64, 58, 0, 52, 4, 277, 52, 7, 16, 6, 96, 51, 44, 19, 18, 83, 28, 167, 43, 12, 26, 64, 375, 14, 0, 50, 3, 40, 91, 2, 147, 65, 9, 88, 43, 24, 14*, 65, 0, 53, 48*, 6, 54, 21, 87, 145, 152, 20, 179, 35, 40, 74, 26, 44, 2, 1, 2, 2, 9, 78, 132, 83, 78, 14, 19, 19, 45, 103, 30, 0, 4, 1, 115, 3, 37, 15, 1, 36, 37, 55, 17, 42, 47, 93, 30, 31, 13*, 89, 11, 7, 4, 39, 51, 79, 4, 33, 68, 14, 62, 3, 213, 8, 153*, 100, 7, 24, 1, 67, 75, 50, 6, 5, 13, 112, 4, 2, 0, 47, 0, 4, 0, 17, 182, 39, 16, 0, 35, 28, 47, 45, 12, 0, 83, 8, 19, 91, 81, 14, 178, 40, 74, 45, 221, 130, 0, 52, 47, 55, 4, 9, 35, 28, 73, 48, 26, 110, 91, 122, 14, 42, 68, 60, 209, 10, 80*, 29, 1, 191, 1, 202, 5, 72, 11, 115, 86, 34, 6, 23, 0, 0, 8, 36, 33, 400*, 53, 120, 11, 44, 95, 13, 0, 7, 79, 15, 196, 4, 176, 13, 4, 130, 48, 153, 0, 5, 36, 30, 14, 13, 45, 226, 17, 5, 0, 1, 1, 83, 18, 0, 7, 120, 10, 19, 26, 11, 61, 122, 216, 0, 49

Zimbabwe

Andy Flower

59, 1*, 81, 14, 9, 115, 62*, 63, 21, 12, 0, 62*, 26, 50, 10, 156, 14, 8, 37, 35, 7, 63, 6, 58*, 35, 45*, 2, 0, 3, 31, 11, 18, 61, 23, 112, 14, 6, 8, 20, 39, 7, 8, 67, 8, 105*, 2, 6, 65, 83, 44, 100*, 1, 49, 30, 41*, 0, 17*, 60*, 28, 0, 13, 39, 8, 14, 86, 15*, 74, 129, 14, 70*, 113*, 5, 66, 10, 24, 2, 42, 29, 22, 48, 65, 183*, 70, 55, 232*, 79, 73, 23, 51, 83, 45, 8*, 142, 199*, 67, 14*, 28, 114*, 42, 10, 8, 11, 6, 3, 3, 8, 92, 0, 29, 67, 30, 13