

# Web-based supplementary materials for “Bayesian GARCH Modeling of Functional Sports Data”

## 1. Latent factor representation for the basis function coefficients

Following [Montagna et al. \(2012\)](#), the basis function coefficients  $\boldsymbol{\theta}_i = [\theta_{i1}, \dots, \theta_{ip}]^\top$  are modelled via a latent factor model as follows:

$$\theta_{ir} = \sum_{l=1}^k \lambda_{rl} \eta_{il} + \xi_{ir}, \quad (1)$$

for  $r = 1, \dots, p$ , where  $\lambda_{rl}$  are the entries of a  $(p \times k)$  factor loading matrix  $\mathbf{\Lambda}$ , and  $\boldsymbol{\eta}_i = [\eta_{i1}, \dots, \eta_{ik}]^\top$  is a vector of  $k$  latent factors for subject  $i$ . Finally,  $\boldsymbol{\xi}_i = [\xi_{i1}, \dots, \xi_{ip}]^\top$  is a residual vector, independent of all other variables in the model. Although other specifications are possible, we follow standard factor analysis modelling choices and assume:

$$\boldsymbol{\eta}_i \stackrel{iid}{\sim} N_k(\mathbf{0}, \mathbf{I}) \quad (2)$$

and the error terms  $\boldsymbol{\xi}_i$  are assumed to have Normal distribution with diagonal covariance matrix,  $\boldsymbol{\xi}_i \stackrel{iid}{\sim} N_p(\mathbf{0}, \text{diag}(\sigma_1^{-2}, \dots, \sigma_p^{-2}))$ , with  $\sigma_j^{-2} \stackrel{iid}{\sim} Ga(a_\sigma, b_\sigma)$ .

The modelling of the factor loading matrix  $\mathbf{\Lambda}$  is crucial for dimensionality reduction. Following [Montagna et al. \(2012\)](#), we adopt a multiplicative gamma process shrinkage (MGSP) prior:

$$\begin{aligned} \lambda_{rl} | \phi_{rl}^{-1}, \tau_l^{-1} &\stackrel{iid}{\sim} N(0, \phi_{rl}^{-1} \tau_l^{-1}) \quad \text{with} \\ \phi_{rl} &\sim Ga\left(\frac{\nu_\phi}{2}, \frac{\nu_\phi}{2}\right) \quad \tau_l = \prod_{v=1}^h \varpi_v \\ \varpi_1 &\sim Ga(a_1, 1) \quad \varpi_v \sim Ga(a_v, 1) \end{aligned} \quad (3)$$

$\varpi_v, v \geq 1$ , are independent. The MGPS prior was originally proposed in [Bhattacharya and Dunson \(2011\)](#) to induce both local and global shrinkage of the factor loading parameters. Specifically,  $\tau_l$  is a global shrinkage parameter for the  $l$ -th column of the loading matrix  $\mathbf{\Lambda}$ . Under a choice  $a_2 > 1$ ,  $\tau_l$  is stochastically increasing, thus favouring more shrinkage as the column index of  $\mathbf{\Lambda}$  increases. Parameters  $\phi_{rl}$  are local shrinkage parameters for the loadings of the  $l$ -th column of  $\mathbf{\Lambda}$ , and prevent over-shrinking the non-zero loadings elements. [Bhattacharya and Dunson \(2011\)](#) discuss extensively properties and computational benefits of using the MGPS prior for the loadings. Figure 1 displays the  $L_2$  norm of the posterior mean estimates of the columns of the factor loading matrix  $\mathbf{\Lambda}$ . The boxplots show a decay in the resulting norms, as expected by the structure induced by the MGPS prior on  $\mathbf{\Lambda}$ . Therefore, the first few important factors are loaded heavily, and significantly contribute to the estimate of  $\boldsymbol{\theta}_i$ . Although the norms of the remaining factors appear to be equal to zero, none of the factor loadings is exactly equal to zero, but shrunk towards zero by the MGPS prior.

Furthermore, [Bhattacharya and Dunson \(2011\)](#) describe a procedure to adaptively select the number of factors  $k$ . Specifically,  $k$  is tuned as the sampler progresses, with adaptations designed to satisfy the diminishing adaptation condition in Theorem 5 of [Roberts and Rosenthal \(2007\)](#). Whenever adaptation occurs, one counts how many columns of  $\mathbf{\Lambda}$  have all elements in some pre-specified neighbourhood of zero. Intuitively, factors corresponding to such columns have a negligible contribution, therefore these columns of  $\mathbf{\Lambda}$  are discarded and one continues the sampler with a reduced number of factors. Otherwise, if the number of such columns drops to zero, one column is added to the loading matrix. When a factor is added, the new parameters are sampled from their prior distributions. Refer to [Bhattacharya and Dunson \(2011\)](#) and [Montagna et al. \(2012\)](#) for further details. Generally, the resulting number of factors

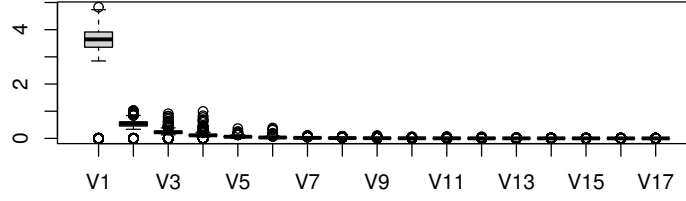


Figure 1: Side-by-side boxplots of the Euclidean norms of the posterior mean estimates of the first 17 columns of the factor loading matrix  $\mathbf{A}$  obtained with a fit of our additive model to the shot put dataset.

is  $k \ll p$ . By combining Eq. (3) in the manuscript with Eq. (1) above one gets:

$$f_i(t) = \sum_{l=1}^k \eta_{il} \Phi_l(t) + r_i(t)$$

where  $\Phi_l(t) = \sum_{r=1}^p \lambda_{rl} b_r(t)$  is a new, unknown non-local basis function learnt from the data, and  $r_i(t)$  a functional error term. Although the number of pre-specified basis functions  $p$  is potentially large, the number of “operative” bases  $k$  is learnt from the data and  $k \ll p$ , thus achieving dimensionality reduction. To provide an example of the estimated learned bases for the shot put data set, we display the first estimated function  $\Phi_1$  in Figure 2. Provided a negative  $\eta_{i1}$  coefficient is estimated, this function explains the overall increasing behaviour in performance over the time span.

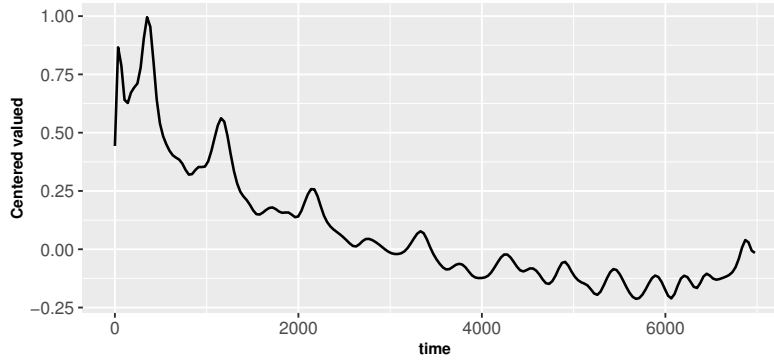


Figure 2: First non-local basis function  $\Phi_1$  learnt from the data of the shot put case study.

We remark that the learnt basis functions are common to all athletes, thus estimated using data from all athletes, and thereby allowing for borrowing of information across athletes. Therefore, if an athlete has sparser observations (e.g., he/she participates to fewer professional competitions) we can still estimate his/her performance trajectory by borrowing from those athletes from whom more information is available.

## 2. Observed error pattern when modelling the seasonal random intercepts as an AR(1)

In a preliminary version or four analyses, we explored an autoregressive dynamic for the individual-specific seasonal means. Specifically, we assumed an autoregressive process of order one, AR(1), to link the subject-specific random intercept at season  $s$ ,  $\mu_{is}$ , to the random intercept in the previous season:

$$\mu_{is} \mid \rho_i, \sigma_\mu^2 \stackrel{iid}{\sim} N(\rho_i \mu_{i(s-1)}, \sigma_\mu^2),$$

where  $\rho_i$  is an autoregressive coefficient, shared across seasons. Although the AR(1) process is straightforward and easily interpretable, there are some modelling drawbacks. First, there is no case study-related argument that justifies why the seasonal mean should depend on previous seasonal means, whereas there exist a consistent literature suggesting that longitudinal datasets present volatility clustering (e.g., [Zhao et al. \(2015\)](#)). By volatility clustering we mean that large changes in the values of a quantity of interest tend to be followed by large changes, whereas small changes (aka fluctuations are smaller in magnitude) tend to be followed by small changes. In relation to our case study, suppose the variance of an athlete's performances during a specific time interval is known, then it is reasonable that this will provide insights on the future variability of his/her measurements. Even further, a history of large volatility for an athlete's performance results provides information about an athlete's potential for large volatility in his/her performance in the future. Second, by looking at the residuals obtained by fitting our additive model (with AR(1) for the random intercepts) to the data, we noticed a particular pattern. Figure 3 displays the error between the model estimates and observations, averaged across athletes and observations in each season. Equivalently, each black point in Figure 3 represents the averaged error (over athletes and their observations) in a given season. These errors fluctuate nicely around zero and with constant variance between seasons 6 and 10, but we notice an increasing trend at the beginning and increased volatility in the errors at later seasons (the spread increases).

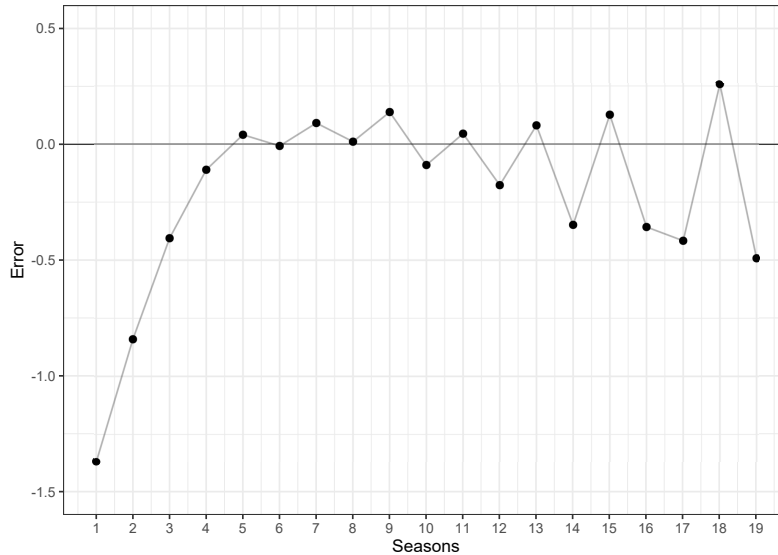


Figure 3: Error (black) between the model estimates and observations, averaged across athletes and observations in each of the 19 sport seasons in the shut put data resulting from modelling the seasonal random intercepts via an AR(1) process.

These findings lead us to consider a normal generalised autoregressive conditional heteroskedastic (GARCH) error for the modelling of the random intercepts, as this formulation helps explaining how variability evolves over periods.

### 3. Posterior inference

In this Section, we discuss in greater detail the full conditional updates anticipated in Section 4.2 of the manuscript and referred to in Algorithm 1 (page 12 of the manuscript).

#### 3.1. Functional model

With reference to the hierarchical model discussed in Section 3 of the manuscript, we get the following updates:

1. Recalling that  $\mathbf{\Lambda}$  is a  $q \times k$  matrix, we update its  $j$ -th row as:

$$\pi(\boldsymbol{\lambda}_j | \dots) \sim N_k((\mathbf{D}_j^{-1} + \sigma_j^{-2} \boldsymbol{\eta}^\top \boldsymbol{\eta})^{-1} \boldsymbol{\eta}^\top \sigma_j^{-2} \boldsymbol{\theta}^{(j)}, (\mathbf{D}_j^{-1} + \sigma_j^{-2} \boldsymbol{\eta}^\top \boldsymbol{\eta})^{-1})$$

where

$$\mathbf{D}_j^{-1} = \text{diag}(\phi_{j1}^{-1} \tau_1^{-1}, \dots, \phi_{jk}^{-1} \tau_k^{-1})$$

$$\boldsymbol{\eta}^\top = (\eta_1, \dots, \eta_k)$$

$$\boldsymbol{\theta}^{(j)} = (\theta_{j1}, \dots, \theta_{jn})$$

2. Concerning the local shrinkage parameters, we get:

$$\pi(\phi_{jh} | \dots) \sim Ga\left(\frac{\nu_\phi + 1}{2}, \frac{\nu_\phi + \tau_h \lambda_{jh}^2}{2}\right)$$

3. The first global shrinkage parameter is updated as follows:

$$\pi(\delta_1 | \dots) \sim Ga\left(a_1 + \frac{qk}{2}, 1 + \frac{1}{2} \sum_{l=1}^k \tau_l^{(1)} \sum_{j=1}^q \phi_{jl} \lambda_{jl}^2\right)$$

4. Whereas for the remaining ones:

$$\pi(\delta_h | \dots) \sim Ga\left(a_h + \frac{q(k-h-1)}{2}, 1 + \frac{1}{2} \sum_{l=1}^k \tau_l^{(h)} \sum_{j=1}^q \phi_{jl} \lambda_{jl}^2\right)$$

with

$$\tau_l^{(h)} = \prod_{t=1, t \neq h}^l \delta_t$$

5. Variances of factor residuals can be sampled as:

$$\pi(\sigma_j^{-2} | \dots) \sim Ga\left(a_\sigma + \frac{n}{2}, b_\sigma + \frac{\sum_{i=1}^n (\boldsymbol{\theta}^{(j)} - \mathbf{\Lambda} \boldsymbol{\eta}_i)^2}{2}\right)$$

6. Latent factors for a specific individual have conditional posterior distribution:

$$\pi(\boldsymbol{\eta}_i | \dots) \sim N(\mathbf{A}_i^{-1} \times \mathbf{B}_i, \mathbf{A}_i^{-1})$$

being

$$\mathbf{A}_i = \mathbf{\Lambda}^\top \mathbf{B}_i^\top (\psi^2 \mathbf{I}_{n_i} + \mathbf{B}_i \boldsymbol{\Sigma} \mathbf{B}_i^\top)^{-1} \mathbf{B}_i \mathbf{\Lambda} + \mathbf{I}_k$$

$$\mathbf{B}_i = \mathbf{\Lambda}^\top \mathbf{B}_i^\top (\psi^2 \mathbf{I}_{n_i} + \mathbf{B}_i \boldsymbol{\Sigma} \mathbf{B}_i^\top)^{-1} \mathbf{y}_i$$

7. Finally, coefficients of the spline representation are sampled from:

$$\begin{aligned}\pi(\boldsymbol{\theta}_i | \dots) &\sim N(\mathbf{C}_i^{-1} \times \mathbf{D}_i, \mathbf{C}_i^{-1}) \\ &\text{being} \\ \mathbf{C}_i &= \psi^2 \mathbf{B}_i^\top \mathbf{B}_i + \boldsymbol{\Sigma}^{-1} \\ \mathbf{D}_i &= \psi^2 \mathbf{B}_i^\top \mathbf{y}_i + \boldsymbol{\Sigma}^{-1} \boldsymbol{\Lambda} \boldsymbol{\eta}_i\end{aligned}$$

### 3.2. Seasonal GARCH model

Concerning the seasonal component, let us start noticing that the likelihood function of the GARCH model is:

$$f(\{\mu_{i,s}\}_{i=1,\dots,n}^{s=1,\dots,S_i} \mid v, h_{i,s}) = \prod_{i=1}^n \prod_{s=1}^{S_i} \frac{1}{\sqrt{2\pi h_{i,s}}} \exp \left\{ -\frac{\overbrace{(\mu_{i,s} - m)^2}^{\zeta_{i,s}^2}}{2h_{i,s}} \right\} \quad (4)$$

whereas the terms in the joint sampling distribution depending on  $\mu_{i,s}$  are:

$$\prod_{i=1}^n \prod_{s=1}^{S_i} \prod_{j \in s} \frac{1}{\sqrt{2\pi\psi^2}} \exp \left\{ -\frac{(y_{i,j}^{(2)} - \mu_{i,s})^2}{2\psi^2} \right\}$$

where

$$y_{i,j}^{(2)} = y_{i,j} - f_i(t_{i,j}) - \mathbf{x}(t_{i,j})\boldsymbol{\beta}$$

is a convenient writing for partial residuals, especially from the implementation point of view. Consequently,  $\mu_{i,s}$  can be updated taking samples from:

$$\begin{aligned}q_\mu(\mu_{i,s} \mid v, y_{i,j}^{(2)}, \psi^2) &\propto \exp \left\{ -\sum_{j \in s} \frac{\mu_{i,s}^2 - 2\mu_{i,s} y_{i,j}^{(2)}}{2\psi^2} \right\} \times \exp \left\{ -\frac{\mu_{i,s}^2 - 2\mu_{i,s} m}{2h_{i,s}} \right\} \\ &\sim N(\hat{\mu}_\mu, \hat{\Sigma}_\mu)\end{aligned}$$

$$\text{Where } \hat{\Sigma}_\mu = \left( \sum_{j \in s} \frac{1}{\psi^2} + \frac{1}{h_{i,s}} \right)^{-1} \text{ and } \hat{\mu}_\mu = \hat{\Sigma}_\mu^{-1} \left( \sum_{j \in s} \frac{y_{i,j}^{(2)}}{\psi^2} + \frac{m}{h_{i,s}} \right)$$

Posterior updates for other parameters are not straightforward in general because of the recursive definition of the conditional variance. As a consequence, we will rely on Metropolis Hastings steps within a Gibbs sampling algorithm. We iteratively generate:

$$\begin{aligned}m^{[j]} &\sim p(m \mid \alpha^{[j-1]}, v^{[j-1]}, \mu_{i,s}) \\ \alpha^{[j]} &\sim p(\alpha \mid m^{[j]}, v^{[j-1]}, \mu_{i,s}) \\ v^{[j]} &\sim p(v \mid m^{[j]}, \alpha^{[j]}, \mu_{i,s})\end{aligned}$$

#### 3.2.1. Update $m$

The posterior distribution for  $m$  is based on the GARCH model under the assumption that the conditional variances  $\{h_{i,s}\}_i^s$  are fixed, known and given by the writing in Eq. (9) of the manuscript. In this case the likelihood function is:

$$f(\{\mu_{i,s}\}_{i=1,\dots,n}^{s=1,\dots,S_i} \mid \boldsymbol{\alpha}, v) = \prod_{i=1}^n \prod_{s=1}^{S_i} \frac{1}{\sqrt{2\pi h_{i,s}}} \exp \left\{ -\frac{(\mu_{i,s} - m)^2}{2h_{i,s}} \right\}$$

Accordingly, the posterior samples for  $m$  will be drawn from:

$$q_m(m \mid \tilde{\alpha}, \tilde{v}) \propto \exp \left\{ - \sum_{i=1}^n \sum_{s=1}^{S_i} \frac{m^2 - 2m\mu_{i,s}}{2h_{i,s}} \right\} \times \exp \left\{ - \frac{m^2 - 2m\mu_{m_0}}{2\Sigma_{m_0}} \right\} \\ \sim N(\hat{\mu}_m, \hat{\Sigma}_m)$$

$$\text{Where } \hat{\Sigma}_m = \left( \sum_{i=1}^n \sum_{s=1}^{S_i} \frac{1}{h_{i,s}} + \frac{1}{\Sigma_{m_0}} \right)^{-1} \text{ and } \hat{\mu}_m = \hat{\Sigma}_m^{-1} \left( \sum_{i=1}^n \sum_{s=1}^{S_i} \frac{\mu_{i,s}}{h_{i,s}} + \frac{\mu_{m_0}}{\Sigma_{m_0}} \right)$$

It is here of capital importance that  $h_{i,s}$  are obtained by means of Eq. (9), using previous realisations the GARCH coefficients from the MH sampler.

### 3.2.2. Update $\alpha$

To generate samples from  $\alpha$  we can not exploit conjugacy, therefore we would like to rely on a normal proposal distribution. Indeed, symmetry of proposal distributions significantly eases computations of acceptance probabilities in Metropolis algorithms. Unfortunately, we are given a non-negativity constraint for the parameter, which forces us to apply a bi-dimensional transformation from the positive quadrant into the real plane: let say  $(\theta_0, \theta_1) = (\log(\alpha_0), \log(\alpha_1))$ . After transformation of the starting parameter, we propose a new sample employing a random walk sampler.

That is, a proposal of the form

$$\theta^* = \theta + \zeta \epsilon$$

where  $\epsilon$  is a multivariate standard normal random vector and  $\zeta$  a covariance matrix, defined according to an Adaptive scaling within the Adaptive Metropolis–Hastings algorithm (ASWAM) by [Haario et al. \(2001\)](#). In this approach, both the covariance matrix of the proposal is adapted to the covariance matrix of the target density and the scale parameter is updated to achieve an average acceptance rate of 0.234 (proven to be optimal in different scenarios).

Provided it will be accepted, the actual parameter will be retrieved applying the inverse transform to  $\theta$ . As discussed above, the acceptance probability will not depend on the proposal distribution. Therefore:

$$\lambda_\alpha = \min \left\{ \frac{p(\theta^* \mid v, m, \mu_{i,s})}{p(\tilde{\theta} \mid v, m, \mu_{i,s})}, 1 \right\} \quad (5)$$

With  $p(\theta \mid v, m, \mu_{i,s})$  the full conditional density of  $\theta$  given the data and the rest of the parameters, which is clearly obtained from  $p(\alpha \mid v, m, \mu_{i,s})$  by change of variables.

### 3.2.3. Update $v$

Similar arguments apply for  $v$ . Samples will be proposed according to a normal distribution centred around  $\tilde{\gamma} = \log(\tilde{v})$ :

$$\gamma^* \sim N(\gamma \mid \tilde{\gamma}, \Sigma_v)$$

Here,  $\Sigma_v$  is defined, for each iteration  $g$  as:

$$\sqrt{\Sigma_v^{(g+1)}} = \zeta^{(g+1)} = \rho(\zeta^{(g)} * w^{(g)}(\lambda_\theta^{(g)} - \bar{\lambda}))$$

to achieve an average optimal acceptance rate of  $\bar{\lambda} = 0.234$  [Haario et al. \(2001\)](#).

Actual samples will be retrieved applying the inverse transform and accepted according to:

$$\lambda_v = \min \left\{ \frac{p(\gamma^* \mid \alpha, m, \mu_{i,s})}{p(\tilde{\gamma} \mid \alpha, m, \mu_{i,s})}, 1 \right\} \quad (6)$$

where again  $p(\gamma \mid \boldsymbol{\alpha}, m, \mu_{i,s})$  is the full conditional density of  $\gamma$  given by change of variables from  $p(v \mid \boldsymbol{\alpha}, m, \mu_{i,s})$

### 3.3. Multiple regression model

On the base of the prior settings and the overall sampling scheme (Eq. (1)-(2)) we get:

1. Regression coefficient, jointly, as:

$$\pi(\boldsymbol{\beta} \mid \dots) \sim N(A^{-1} \times B, A^{-1})$$

being

$$A = \mathbf{x}^\top \mathbf{x} / \psi^2 + \Sigma_{\beta_0}$$

$$B = \Sigma_{\beta_0} \boldsymbol{\beta}_0 + \mathbf{x}^\top \mathbf{y}^{(3)} / \psi^2$$

2. The variance hyperparameter instead is updated as:

$$\pi(\sigma_\beta^{-2} \mid \dots) \sim Ga\left(\frac{N + \nu_\beta}{2}, \frac{\nu_\beta \sigma_\beta^2 + \sum_{i=1}^n \sum_{j=1}^{n_i} (y_{ij} - f_i(t_{ij}) - \mu_{is} - \mathbf{x}\boldsymbol{\beta})^2}{2}\right)$$

### 3.4. Error term

Finally, the full conditional posterior distribution for the model variance is:

$$\pi(\psi^{-2} \mid \dots) \sim Ga\left(\frac{N + \nu_\psi}{2}, \frac{\nu_\psi \sigma_\beta^2 + \sum_{i=1}^n \sum_{j=1}^{n_i} \epsilon_{i,j}^2}{2}\right)$$

being

$$\epsilon_{i,j} = y_{i,j} - f_i(t_{i,j}) - \mu_{i,s} - \mathbf{x}\boldsymbol{\beta}$$

#### 4. Posterior convergence diagnostics

In this Section, we report convergence diagnostics for model  $M_1$  fit to shot put data set (Section 5 of the manuscript). Posterior sampling is done as described in Section 4 of the manuscript and, in particular, we ran a chain of 20,000 iterations with a burn-in period of 50% and a thinning of 5. Thus, posterior sampling leverages on 2000 samples. The execution time is about 5 hours on a personal computer with the following characteristics: Processor Intel(R) Core(TM) i7-10875H CPU @ 2.30GHz with 32.0 GB of installed RAM.

We rely on two diagnostics: 1) traceplots, and 2) the effective sample size ( $ESS$ ), which is defined as the sample size adjusted for the autocorrelation in the posterior samples:

$$ESS = \frac{G}{1 + 2 \sum_{g=1}^G \widehat{Corr}(\theta^{(0)}, \theta^{(g)})},$$

where  $\widehat{Corr}(\theta^{(0)}, \theta^{(g)})$  is the autocorrelation at lag  $g$  of posterior draws of a generic model parameter  $\theta$ , and  $G$  is the number of retained posterior samples. Refer to, e.g., [Gelman and Rubin \(1992\)](#) and [Gelman et al. \(1995\)](#) for definitions.

Figure 4 provides trace plots for a selection of model parameters, and Table 1 reports their  $ESS$ . In general, we notice how sampling is sufficiently close to Normal independent sampling. A special remark is in order concerning the first GARCH coefficient:  $\nu$  is positive and, as such, is sampled from a truncated normal. Accordingly, its trace plot is reasonably well behaved. Similar conclusions about convergence are drawn by looking at other model parameters (e.g., other athlete- and season-specific random intercepts and athlete-specific basis function coefficients), or by investigating convergence for different model specifications (models  $M_1^{(2)}$ - $M_6$  in the manuscript). The only parameter for which we have evidence of slower convergence is the GARCH parameter  $\alpha_1$ . Notice that the GARCH parameters are non conjugate and that they require a MH step within Gibbs. Coupled with the constraints imposed on these quantities, this might have affected the acceptance probabilities of the relative chains.

Table 1: Effective sample sizes for a selection of model parameters estimated under model  $M_1$  (Section 5 of the manuscript).

|     | $\beta_{Environment}$ | $\nu$ | $\alpha_1$ | $\zeta_{72,4}$ | $m$ | $\mu_{3,6}$ | $\theta_{39,2}$ | $\psi$ |
|-----|-----------------------|-------|------------|----------------|-----|-------------|-----------------|--------|
| ESS | 1820                  | 1332  | 241        | 1477           | 402 | 1839        | 1897            | 654    |

Finally, let us recall that the random intercept model is  $\mu_{is} \mid m, h_{is} \stackrel{iid}{\sim} N(m, h_{is})$ , where  $m$  is the overall mean parameters and  $h_{is}$ 's are GARCH errors. As stated in the manuscript (Section 3.2), an informative Normal prior is placed on  $m$ . In particular, we take an empirical Bayes approach by setting the hyperparameters of the Normal prior on  $\mu$  such that, a priori,  $\mathbb{E}[m]$  is the sample mean of the mean result by season, and  $Var(m)$  its sample variance (Table 4 of the manuscript). Our results indicate that a non-informative specification for  $m$  leads to a lower LPLM than that achieved with an informative specification on  $m$  (see the LPLM results for  $M_1$  and  $M_1^2$  in Table 3 of the manuscript), all other things being equal. Further, a non-informative prior makes convergence of the chain much slower. Specifically, we run the chain for 30,000 iterations with a large burn-in, but yet we did not have evidence of convergence. The last draws for the model parameters were starting to converge around similar values to those resulting from the informative case, however acceptable ESSs were reached only after a number of iterations that was too demanding computationally.

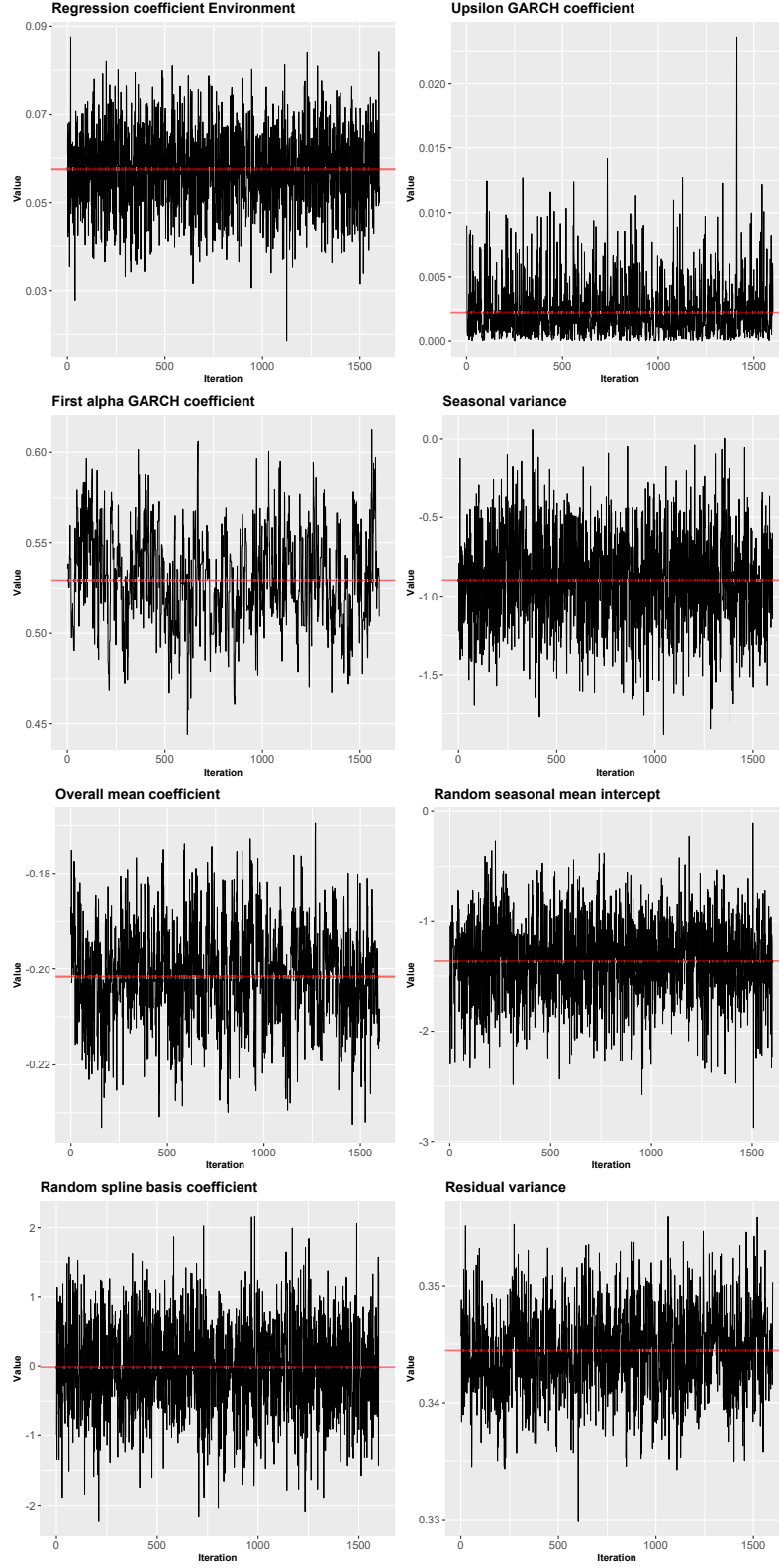


Figure 4: Trace plots for a selection of model parameters estimated under model  $M_1$  (Section 5 of the manuscript). Panels refer to, respectively,  $\beta_{sex}$ ,  $\beta_{env}$ ,  $v$ ,  $\alpha_1$ ,  $m$ ,  $\mu_{3,6}$ ,  $\theta_{39,2}$ , and  $\psi$ .

## 5. Explained variance

Results in Section 5 of the manuscript illustrate that the seasonal random intercepts are highly informative when explaining the variability in the data. In this Web Section, we quantify the proportion of variance explained by the random intercepts by means of the variance decomposition formula. Specifically, let us construct the partial residuals:

$$y_{ij}^{(2)} = y_{ij} - f_i(t_{ij}) - \mathbf{x}_i(t_{ij})\boldsymbol{\beta},$$

where the  $y_{ij}^{(2)}$  is the residual (at time  $j$ ) quantifying any variability in the data not explained by the functional ( $f_i$ ) and regressive components ( $\mathbf{x}_i(t_{ij})\boldsymbol{\beta}$ ) of our model.

Following the decomposition of variance formula in “within group” and “between groups” variance, we follow the reasoning in Hoff (2009) (Chapter 8) and interpret our seasons as the multiple groups. Then, the between-groups variance is the variance of the error term of our model,  $\psi^2$  (Eq. (1) in the manuscript) whereas the within-group variance is the variance of the random intercepts  $\mu_{is}$ ’s (Eq. (8) in the manuscript). Thus, we quantify the proportion of variance explained by the GARCH component of our model, for each athlete  $i$  and at each iteration  $g$  of our MCMC sampler, as:

$$R_i^{2(g)} = \frac{\text{Var}(\mu_{is}^{(g)})}{\text{Var}(\mu_{is}^{(g)}) + \psi^{(g)}}$$

Averaging over iterations and athletes we get an estimate of the desired quantity, aka the explained variance coefficient, taking value 0.564. We conclude that, indeed, the majority of the variability occurs between seasons, that is in jumps across seasons.

Our additive model formulation does not include an interaction effect between the seasonal random intercept and the functional component, nor an interaction between the regressive component and the functional/GARCH components. Therefore, we can not provide a model-based evaluation of whether and how covariates interact with intra-season and inter-seasons variability. Specifically, we can not provide an assessment of whether, e.g., younger players have more intra- and/or inter-seasonal variability, or how is this variability changing across covariates levels. Heuristically, however, we can compute the explained variance coefficient for specific covariate levels of interest. For example, we can compute the  $R^2$  (averaging across iterations) for athletes belonging to a particular age group (thus, averaging only over athletes belonging to a specific age group), or compare the  $R^2$ ’s between the two sexes. Tables 2 and 3 report some results of this evaluation. Table 3 seem to suggest that younger athletes display more inter-season variability than more mature ones. This might be due to the faster performance progress that occurs in the early stages of an athlete’s career.

Table 2:  $R^2$  (averaged across iterations) for male and female athletes.

| Gender | $R^2$ |
|--------|-------|
| Male   | 0.56  |
| Female | 0.60  |

Table 3:  $R^2$  (averaged across iterations) for different age groups based on the quantiles of the age distribution in the dataset.

| Age       | $R^2$ |
|-----------|-------|
| [13 – 22) | 0.57  |
| [22 – 24) | 0.32  |
| [24 – 47) | 0.19  |

## 6. Model comparison across additive components

In this Section, we investigate how the three different components of our additive model, namely, the functional, GARCH and regressive components (Eq. (1)-(7) in the manuscript), contribute to the overall model fit. To this end, consider the model:

$$y_{ij} = g_i(t_{ij}) + \epsilon_{ij}$$

with  $\epsilon_{ij} \stackrel{iid}{\sim} N(0, \psi^2)$  independent errors, and  $g_i(t_{ij})$  is modelled as, respectively:

$$g_i(t_{ij}) = \sum_{r=1}^p \theta_{ir} b_r(t_{ij}) + \mu_{is} + \mathbf{x}_i(t_{ij})\boldsymbol{\beta} \quad (7)$$

$$g_i(t_{ij}) = \sum_{r=1}^p \theta_{ir} b_r(t_{ij}) \quad (8)$$

$$g_i(t_{ij}) = \mu_{is} \quad (9)$$

Eq. (7) represents the full model, complete of all three additive components discussed in the paper. In the formulation given by Eq. (8), the model reduces to the functional latent factor regression model in [Montagna et al. \(2012\)](#). In the formulation given by Eq. (9), the model reduces to the GARCH component, namely, a random intercept model where  $\mu_{is}$  is an athlete- and season-specific random intercept modelled as

$$\mu_{is} = m + \zeta_{i,s}, \quad \text{with} \quad \zeta_{i,s} \stackrel{iid}{\sim} N(0, h_{i,s}) \quad (10)$$

The resulting estimated athlete-specific function  $g_i$  is a step function where, in each season, performance is estimated as the average performance of that season.

We compare formulations (7)-(9), hereafter called “complete”, “functional”, and “GARCH”, respectively, in terms of predictive RMSE and training root mean squared error (RMSE: squared discrepancy between measured and estimated ( $\hat{g}_i$ ) performance, averaged over all athletes and training data and squared rooted):

$$RMSE_{training} = \frac{1}{N} \sum_{i=1}^N \sqrt{\frac{1}{n_i} \sum_j (y_{ij} - \hat{g}_i(t_{ij}))^2},$$

where  $N$  denotes the number of athletes belonging to the training set, and  $n_i$  the number of observations per athlete. The predictive RMSE is computed on held-out data, as described in our predictive tests (Section 5.3 of the manuscript). While the training RMSE is an indicator of model fit, the predictive RMSE is indicative of the predictive accuracy of the models on the test data. Results reported in Table 4 support the complete formulation of our model (Eq. (7)) over alternative formulations.

Table 4: Comparison of formulations given by Eq. (7)-(9) by means of the root mean squared error (RMSE). The second column reports the training RMSE whereas the third column reports the averaged RMSE computed on held-out data from central seasons (Average RMSE Cent). The fourth column reports the same quantity for held-out data from terminal seasons of athletes' careers (Average RMSE Fin).

| Model      | Train RMSE | Average RMSE Cent | Average RMSE Fin |
|------------|------------|-------------------|------------------|
| Complete   | 334.28     | 0.304             | 0.949            |
| GARCH      | 347.67     | 0.7645            | 0.983            |
| Functional | 600.09     | 0.514             | 0.989            |

## 7. Estimated regression coefficients for models $M_2$ , $M_5$ , and $M_6$

Table 5: Posterior mean estimate of the regression coefficients for model  $M_2$  (Table 3 in the Manuscript), together with the standard deviation of their estimate, effective sample size (ESS) with respect to 1600 retained samples, and 95% posterior credible bounds.

| Coeff     | Mean       | Sd        | ESS  | 2.5%       | 97.5%     |
|-----------|------------|-----------|------|------------|-----------|
| $\beta_1$ | -5.58 e-03 | 0.0263    | 223  | -0.0585    | 2.00 e-03 |
| $\beta_2$ | 3.34 e-04  | 1.16 e-03 | 107  | -2.60 e-03 | 2.00 e-03 |
| $\beta_3$ | 0.0570     | 9.09 e-03 | 1600 | 0.0395     | 0.0743    |

Table 6: Posterior mean estimate of the regression coefficients for model  $M_5$  (Table 3 in the Manuscript), together with the standard deviation of their estimate, effective sample size (ESS) with respect to 600 retained samples, and 95% posterior credible bounds.

| Coeff     | Mean      | Sd        | ESS | 2.5%     | 97.5%    |
|-----------|-----------|-----------|-----|----------|----------|
| $\beta_1$ | -0.120    | 0.0285    | 77  | -0.178   | -0.0690  |
| $\beta_2$ | 6.71 e-03 | 1.11 e-03 | 64  | 4.40 e-3 | 8.10 e-3 |
| $\beta_3$ | 0.0373    | 0.0108    | 600 | 0.0168   | 0.0580   |
| $\beta_4$ | -0.116    | 0.0761    | 70  | -0.267   | -0.0365  |

Table 7: Posterior mean estimate of the regression coefficients for model  $M_6$  (Table 3 in the Manuscript), together with the standard deviation of their estimate, effective sample size (ESS) with respect to 600 retained samples, and 95% posterior credible bounds.

| Coeff     | Mean       | Sd        | ESS | 2.5%      | 97.5%     |
|-----------|------------|-----------|-----|-----------|-----------|
| $\beta_1$ | -0.0104    | 0.0250    | 74  | -0.0617   | 0.0362    |
| $\beta_2$ | -4.60 e-04 | 1.14 e-03 | 67  | 2.80 e-03 | 1.90 e-03 |
| $\beta_3$ | 0.0578     | 9.37 e-03 | 600 | 0.0408    | 0.0756    |
| $\beta_4$ | -0.0103    | 0.0638    | 91  | -0.134    | 0.115     |

## 8. Prediction performance and model validation

Figure 5 and 6 show the estimated performance trajectories for three randomly selected test set athletes in the mid-career and end-of-career performance prediction analyses presented in Section 5.3 of the manuscript.

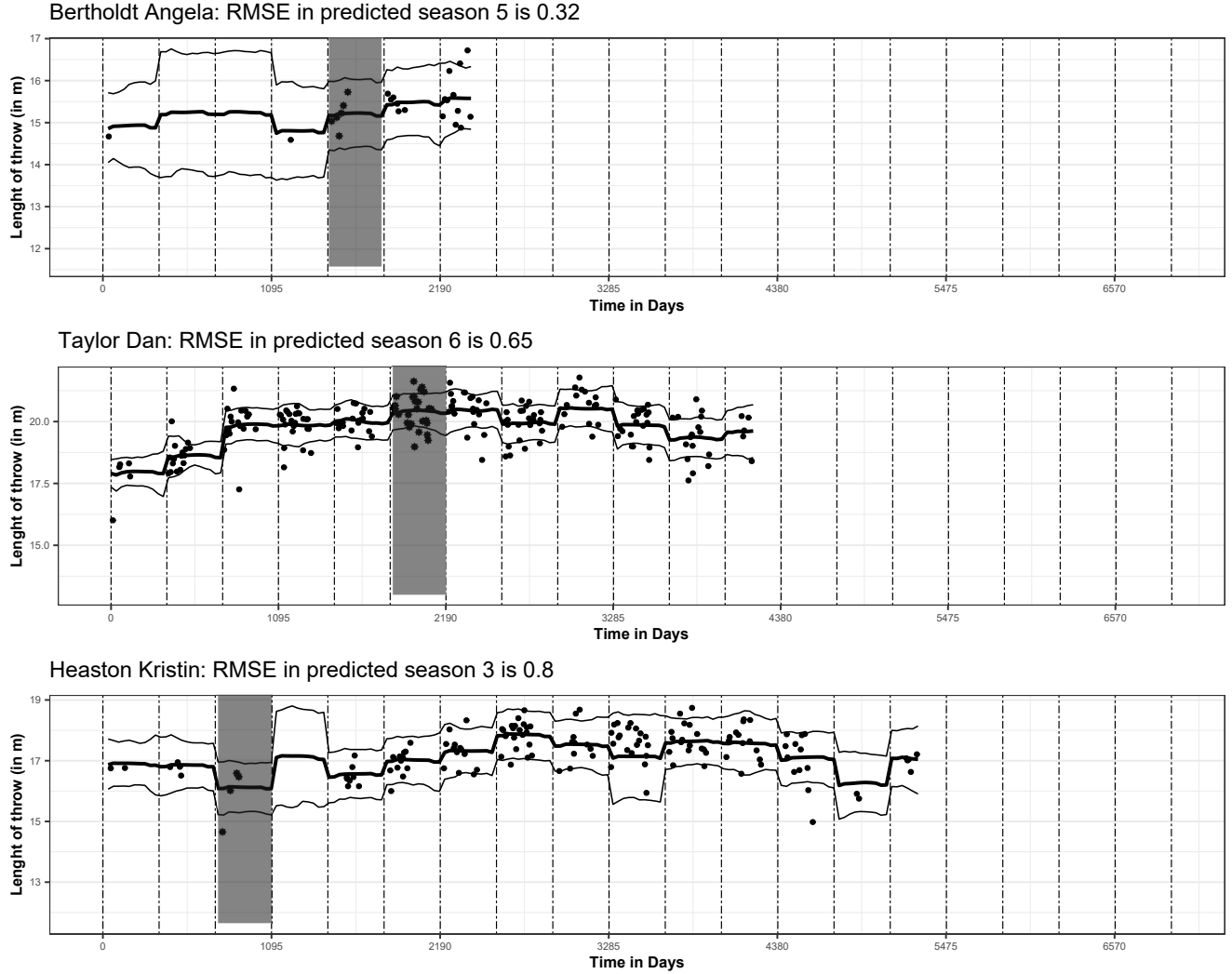
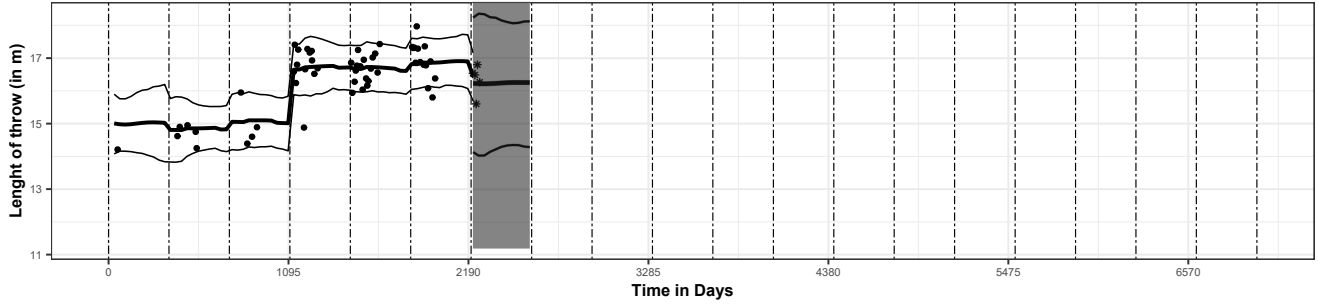
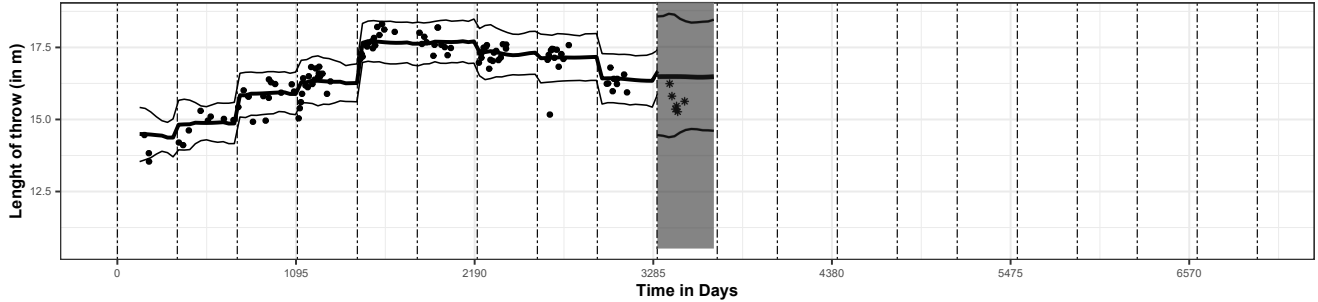


Figure 5: Estimated performance trajectories (and 95% credible bands) for three randomly selected test set athletes in the mid-career performance prediction analysis (Section 5.3 of the manuscript). Held out data in each panel are plotted via asterisks.

Shump Karen: RMSE in predicted season 7 is 0.42



Labonte Julie: RMSE in predicted season 10 is 0.91



Samuels Kelsey: RMSE in predicted season 10 is 0.29

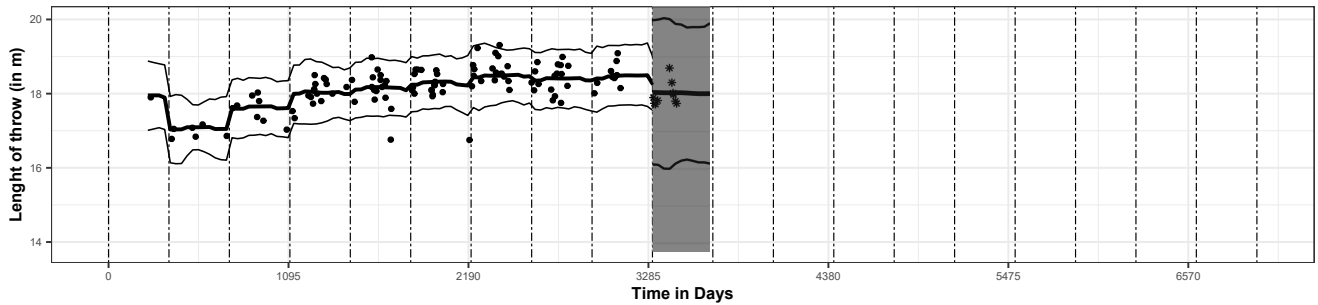


Figure 6: Estimated performance trajectories (and 95% credible bands) for three randomly selected test set athletes in the end-of-career performance prediction analysis (Section 5.3 of the manuscript). Held out data in each panel are plotted via asterisks.

## References

- Bhattacharya A, Dunson D (2011) Sparse bayesian infinite factor models. *Biometrika* 98:291–306, DOI 10.2307/23076151
- Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statistical science* 7(4):457–472
- Gelman A, Carlin JB, Stern HS, Rubin DB (1995) *Bayesian data analysis*. Chapman and Hall/CRC
- Haario H, Saksman E, Tamminen J (2001) An adaptive metropolis algorithm. *Bernoulli* 7:223–242
- Hoff P (2009) *A First Course in Bayesian Statistical Methods*. DOI 10.1007/978-0-387-92407-6
- Montagna S, Tokdar S, Neelon B, Dunson D (2012) Bayesian latent factor regression for functional and longitudinal data. *Biometrics* 68:1064–1073, DOI 10.2307/41806025
- Roberts GO, Rosenthal JB (2007) Coupling and ergodicity of adaptive mcmc. *Statistical science* 44:458–475

Zhao H, Huffer F, Niu XF (2015) Time-varying coefficient models with arma-garch structures for longitudinal data analysis. *Journal of Applied Statistics* 42(2):309–326