

Research Article

Prediction Accuracy in Multivariate Repeated-Measures Bayesian Forecasting Models with Examples Drawn from Research on Sleep and Circadian Rhythms

Clark Kogan,¹ Leonid Kalachev,² and Hans P. A. Van Dongen^{1,3}

¹*Sleep and Performance Research Center, Washington State University, Spokane, WA 99210, USA*

²*Department of Mathematical Sciences, University of Montana, Missoula, MT 59812, USA*

³*Elson S. Floyd College of Medicine, Washington State University, Spokane, WA 99210, USA*

Correspondence should be addressed to Hans P. A. Van Dongen; hvd@wsu.edu

Received 23 May 2015; Accepted 27 August 2015

Academic Editor: Chung-Min Liao

Copyright © 2016 Clark Kogan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In study designs with repeated measures for multiple subjects, population models capturing within- and between-subjects variances enable efficient individualized prediction of outcome measures (response variables) by incorporating individuals response data through Bayesian forecasting. When measurement constraints preclude reasonable levels of prediction accuracy, additional (secondary) response variables measured alongside the primary response may help to increase prediction accuracy. We investigate this for the case of substantial between-subjects correlation between primary and secondary response variables, assuming negligible within-subjects correlation. We show how to determine the accuracy of primary response predictions as a function of secondary response observations. Given measurement costs for primary and secondary variables, we determine the number of observations that produces, with minimal cost, a fixed average prediction accuracy for a model of subject means. We illustrate this with estimation of subject-specific sleep parameters using polysomnography and wrist actigraphy. We also consider prediction accuracy in an example time-dependent, linear model and derive equations for the optimal timing of measurements to achieve, on average, the best prediction accuracy. Finally, we examine an example involving a circadian rhythm model and show numerically that secondary variables can improve individualized predictions in this time-dependent nonlinear model as well.

1. Introduction

Significant steps forward in the analysis of repeated-measures data were made with the introduction of linear and nonlinear mixed-effects models [1–3], which distinguish within-subjects variance (from multiple measurements in each subject) versus between-subjects variance (from multiple subjects being measured). Distinguishing these types of variance can also be thought of as explicitly modeling random error in the data. This can be useful in understanding how different individuals are from one another as compared to how different multiple measurements are for given individuals. In research on sleep and sleepiness, for example, breakthroughs made possible by mixed-effects models include elucidation of the dose-response effects of sustained sleep restriction on

sleep architecture and neurobehavioral impairment [4, 5] and demonstration of the trait characteristics of individual differences in vulnerability to sleep loss [6]. In recent years, mixed-effects model approaches to statistical regression and analysis of variance have become widely available in statistical software packages. They are nowadays the methodology of choice for many repeated-measures investigations in sleep research and other fields of study.

A further advance was the introduction of a model individualization technique called Bayesian posterior distribution estimation or Bayesian forecasting. This technique was first used in sleep research to overcome a shortcoming of biomathematical models of fatigue and performance. Existing models did not account for individual differences in sleep regulation and vulnerability to sleep loss and therefore

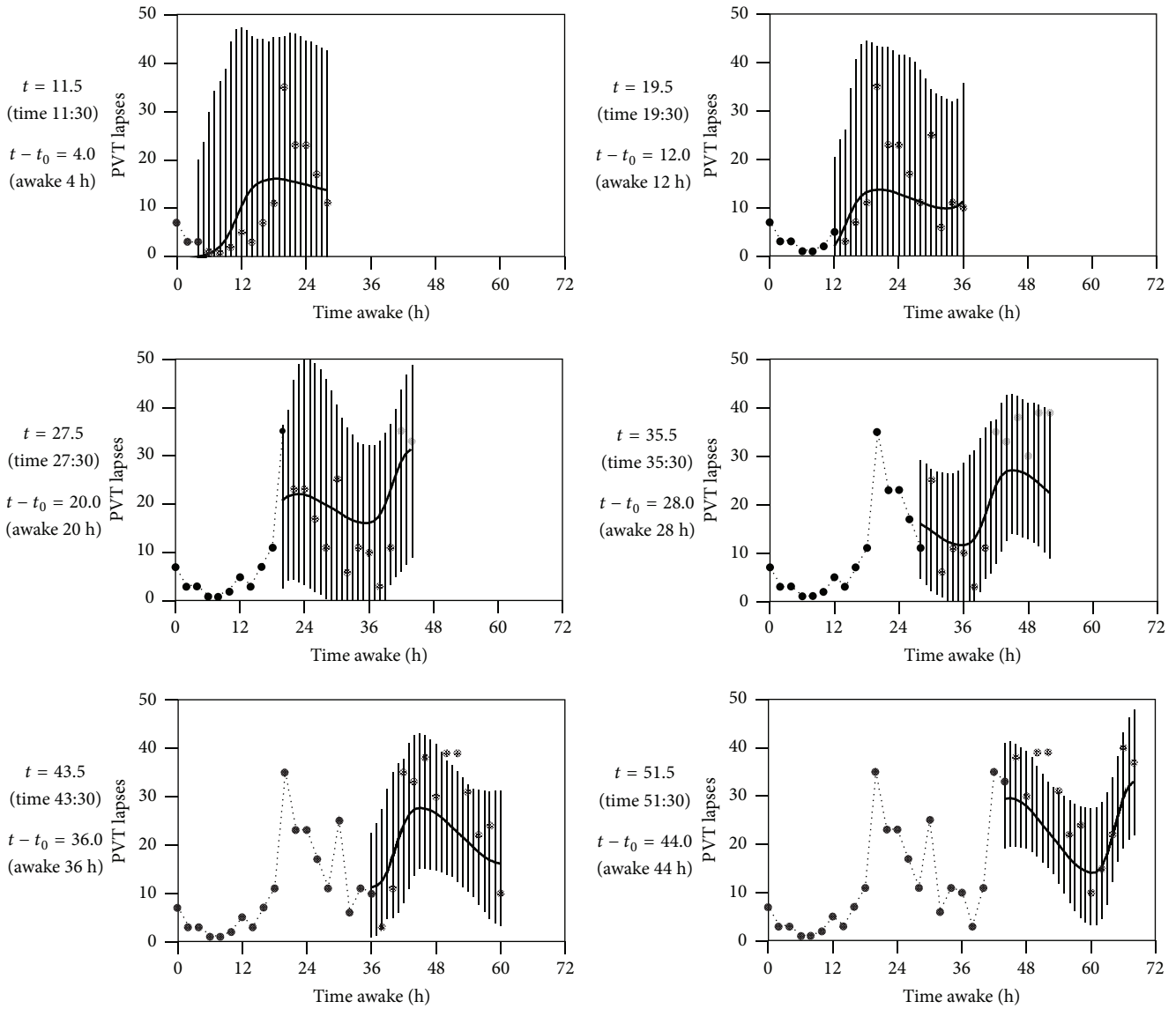


FIGURE 1: Lapses of attention on a psychomotor vigilance test (PVT) for a subject in a study involving 88 h of total sleep deprivation under controlled laboratory conditions. In each of the six plots, different amounts of subject data are assumed known (black dots), and the Bayesian forecasting procedure is applied to the known data to construct predictions of PVT number of lapses for a 24 h interval immediately following the most recent collected data point at time t . For the 24 h interval, 95% prediction intervals (vertical bars) are shown, along with the remainder of the data for comparison (gray dots). The beginning of the 88 h sleep deprivation period, t_0 , was at 07:30. Graphs taken from [7] with permission.

did not accurately predict performance for given individuals. Bayesian forecasting addressed this shortcoming by utilizing the separation of within- and between-subjects variance in model parameters as enabled by mixed-effects modeling [2]. In Bayesian forecasting, the between-subjects variance of model parameters serves as Bayesian prior information. Measurements from a new individual, not previously studied, are combined with the prior information to efficiently derive model parameters tailored to the new individual, thereby yielding a subject-specific mathematical model [2, 7, 8]. As a bonus, the Bayesian forecasting technique also yields

quantitative estimates of the accuracy of individualized predictions made with the subject-specific mathematical model [9].

In a published example, Bayesian forecasting was implemented for the two-process model of sleep regulation [10, 11] to predict performance impairment of selected individuals undergoing a period of total sleep deprivation (see Figure 1). Comparisons with the individuals' actual data revealed that the model parameters converged efficiently to those that best characterized each individual, and the response predictions were significantly more accurate than could have been

achieved with the original, nonindividualized two-process model [7].

In Bayesian forecasting, individualized parameter estimates are derived from the posterior distributions of the parameters in question after combining prior distributions with the measurement(s) from the individual, and individualized model predictions for future outcomes or responses are obtained following estimation of the posterior distribution of the expected responses at given times. A variety of methods are available for obtaining parameter estimates and response predictions once posterior distributions have been estimated. We employ the Bayesian mean squared error (BMSE) [12] and make use of the Bayesian minimum mean squared error (MMSE) estimator [13, 14], which produces unbiased point estimates that minimize the BMSE [15]. The BMSE of parameter estimates and response predictions is dependent on the amount of data available for the individual at hand and the magnitude of between-subjects variance captured in the Bayesian prior distributions.

In cases where between-subjects variance is relatively large, such as performance responses to sleep loss [6, 16], measurement data for the individual at hand are more critical for making accurate individualized response predictions. Sparseness of such measurement data (e.g., due to practical or cost-based limitations) can result in unacceptably low levels of accuracy. For example, Bayesian forecasting could be used to develop a drowsy driver warning system based on a mathematical model of fatigue and performance [17, 18] calibrated to predict lateral lane deviation, using camera-based measurements of lane position to individualize the model for the driver. However, when lane markers are covered with snow and cameras are unable to determine vehicle position relative to the lane, the individualization effort becomes less effective.

To address this limitation, we consider the use of secondary variables to increase data availability, boost response prediction accuracy, and/or reduce data collection costs for individualized response prediction. For example, in the case of a drowsy driver warning system, camera-based measurements of lateral lane deviation serving as the primary response variable could be augmented with in-car secondary variables such as steering wheel variability or driver eyelid closure assessments. However, individualization of predictions based on two or more measurement variables would only be straightforward if individual differences in responses on these variables are equivalent (cf. [19]). Generally, this is not what the evidence shows. As a case in point, trait individual differences in vulnerability to performance impairment due to sleep loss vary considerably across outcome variables [6, 20], such that the most vulnerable individuals based on one variable are not necessarily also the most vulnerable individuals based on another variable. Therefore, when considering two or more response variables as the basis for individualized prediction, it is essential to account statistically for the level of congruence between the response variables.

Here we develop a multivariate statistical framework for individualized prediction of sleep or performance variables, based on Bayesian forecasting with measurements of a

primary response variable, augmented with one or more measurements of secondary response variables. The response variables are assumed to follow equivalent dynamics over time, such that they can be described by the same model framework after appropriate scaling. This is a reasonable assumption in the case of, for example, models describing sleep variables measured repeatedly across multiple nights [21] and models describing performance changes over time in response to sleep loss [22]. We make use of multivariate Bayesian prior distributions of the primary and secondary variables, assumed to have been assessed in advance by means of mixed-effects modeling [2] or other suitable techniques. The between-subjects correlation(s) between the primary and secondary variables, used here to account for the level of congruence between the response variables, is assumed to have been estimated as part of the covariance matrix of the multivariate Bayesian prior distributions. The between-subjects correlation(s) are assumed to be at least moderately strong, lest the secondary variables contain essentially no information about the primary response variable that rises above the level of measurement noise.

Our focus in this paper is on prediction accuracy in the multivariate Bayesian forecasting technique. We develop the technique by first considering the details of making individualized response predictions and estimating their accuracy for a simple univariate intercept model. To demonstrate how secondary, correlated responses can be used to make more accurate individualized predictions, we expand the intercept model to include both primary and secondary responses. Assuming fixed costs of data collection for each of the responses, we show how to optimize data collection given a desired level of accuracy for predictions of the primary response variable.

We then consider a linear approximation of the homeostatic component of the two-process model [11] and derive a closed form equation of the BMSE to quantify prediction accuracy in this time-dependent model. For this example, we study the problem of optimizing the timing of measurements in order to enhance the individualized prediction accuracy most efficiently. Finally, we consider more complicated bivariate models, both linear and nonlinear, for which the BMSE cannot be determined in closed form. For these models, we describe a process of numerically assessing the BMSE for individualized prediction based on a primary response variable in Bayesian forecasting augmented with measurements of a secondary response.

2. Subject-Specific Bayesian Models

First we discuss a modeling framework for Bayesian forecasting. Consider a response variable y , dependent on a subject-specific trait parameter θ . Let i be an index for individual, ranging from 1 to N , and let j be an index for observations ordered by time, nested within individual. Suppose that the response variable can be modeled by

$$y_{ij} = f(t_{ij}, \theta_i) + \epsilon_{ij}, \quad (1)$$

where $f(t_{ij}, \theta_i)$ represents the model function, t_{ij} represents a fixed measurement time, θ_i represents a random subject-specific parameter, and ϵ_{ij} represents additive measurement error. We assume that the distributions for θ_i and ϵ_{ij} are known. Equation (1) and the distributions of θ_i and ϵ_{ij} constitute a population model.

Limiting our focus to a particular individual, we may remove the subscript i and model the subject's responses as

$$y_j = f_j + \epsilon_j, \quad (2)$$

where f_j is used to denote $f(t_j, \theta)$.

Suppose that a total of m responses (y_1, \dots, y_m) have been measured for the individual at hand, and let j^* denote the index of a response y_{j^*} at some future time t_{j^*} . We consider a prediction (estimator) of the expected response $E[y_{j^*} | \theta] = f_{j^*}$, which we denote as \hat{f}_{j^*} . Our interest is in constructing \hat{f}_{j^*} such that the expected accuracy for an arbitrary, given individual from the population is minimized.

We define the accuracy using the squared error $(f_{j^*} - \hat{f}_{j^*})^2$. The expected accuracy, which is referred to as the Bayesian mean squared error (BMSE), is thus given by

$$\mathbf{M}_{\hat{f}_{j^*}} \equiv \text{BMSE}(\hat{f}_{j^*}) \equiv E[(f_{j^*} - \hat{f}_{j^*})^2], \quad (3)$$

where the expectation is taken with respect to the marginal probability density function (pdf) of y_1, \dots, y_m and θ .

We refer to the prediction that minimizes the BMSE as a minimum mean squared error (MMSE) prediction. After observing data from a particular individual, y_1, \dots, y_m , and constructing the MMSE prediction \hat{f}_{j^*} , we seek to assess the expected accuracy of this particular prediction. This can be done with confidence intervals on f_{j^*} , obtained from quantiles of the posterior distribution of $\theta | y_1, \dots, y_m$. In the following sections, we describe specific types of models and investigate the BMSE and the MMSE that minimizes it.

3. Univariate Random Intercept Model

We consider a random intercept model obtained from (2) by letting $f_j = \theta$:

$$y_j = \theta + \epsilon_j, \quad (4)$$

$$\theta \sim \mathcal{N}(\mu, \delta^2), \quad (5)$$

$$\epsilon_j \sim \mathcal{N}(0, \sigma^2), \quad (6)$$

where $j = 1, \dots, m$. For a particular individual, the expected response at time t_{j^*} is

$$E[y_{j^*} | \theta] = f_{j^*} = \theta. \quad (7)$$

It follows that the MMSE prediction of $E[y_{j^*} | \theta]$ is

$$\hat{f}_{j^*} = \hat{\theta}, \quad (8)$$

where $\hat{\theta}$ is the MMSE of θ . The estimator $\hat{\theta}$ (and therefore \hat{f}_{j^*}) is given by [23]:

$$\hat{\theta} = v\bar{y} + (1-v)\mu, \quad (9)$$

where \bar{y} is the sample mean of the measured responses and

$$v = \frac{\delta^2}{\delta^2 + \sigma^2/m}. \quad (10)$$

The variance of the posterior distribution of $\theta | y_1, \dots, y_m$ (and therefore $\hat{f}_{j^*} | y_1, \dots, y_m$) is [23]:

$$\text{Var}(\theta | y_1, \dots, y_m) = \frac{\sigma^2}{m} \left(\frac{\delta^2}{\delta^2 + \sigma^2/m} \right). \quad (11)$$

Furthermore, the BMSE of $\hat{\theta}$ (and therefore \hat{f}_{j^*}) is given by [23]:

$$\mathbf{M}_{\hat{\theta}} = \text{Var}(\theta | y_1, \dots, y_m). \quad (12)$$

The MMSE prediction \hat{f}_{j^*} (given by (8) and (9)) represents a trade-off between knowledge about the population and knowledge about the subject at hand. This trade-off is embodied by the weighting factor v in (10). When no data are available for the subject at hand, $v = 0$, resulting in a prediction made at $\hat{f}_{j^*} = \mu$, where μ is, in this case, representative of the population mean $E[\theta]$ as well as the population mean response $E[y_{j^*}]$:

$$E[y_{j^*}] = E[\theta] = \mu. \quad (13)$$

As we begin to collect *subject-specific* data, the weighting factor moves towards the value $v = 1$, resulting in a prediction that converges to the subject-specific sample mean as data collection continues.

Likewise, when no data are available for the subject at hand, the expected accuracy of the prediction is $\mathbf{M}_{\hat{f}_{j^*}} = \delta^2$. The term δ^2 is, in this case, representative of the population variance as well as the variance in mean response over the population:

$$\begin{aligned} \text{Var}[E[y_{j^*} | \theta]] &= E[(E[y_{j^*} | \theta] - E[E[y_{j^*} | \theta]])^2] \\ &= E[(E[y_{j^*} | \theta] - \mu)^2] = E[(\theta - \mu)^2] \\ &= \delta^2, \end{aligned} \quad (14)$$

where the expectation $E[y_{j^*} | \theta]$ is evaluated with respect to the conditional distribution of $y_{j^*} | \theta$, and all other expectations are evaluated with respect to the marginal distribution of y_{j^*} . Per (12), as we begin to collect subject-specific data, the expected accuracy of the prediction improves (i.e., the BMSE decreases, where smaller is better).

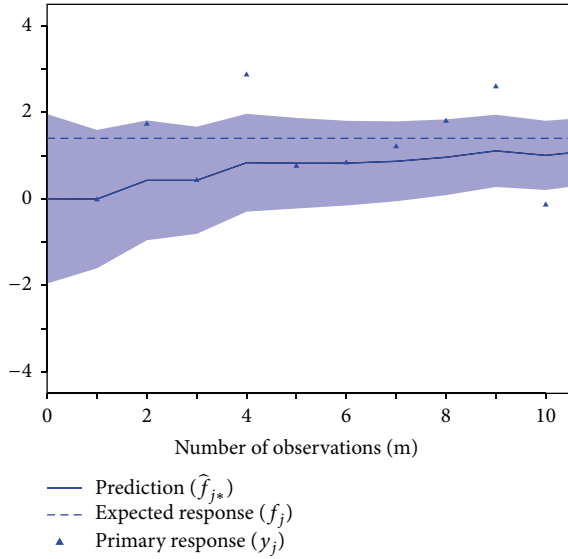


FIGURE 2: MMSE predictions of the expected response \hat{f}_{j^*} from the univariate random intercept model specified in (4), and corresponding 95% confidence intervals, determined by incorporating simulated observations y_j iteratively. The individualized Bayesian forecasting predictions begin at the population mean expected response $\mu = 0$ when no individual data are used and converge to the subject's expected individualized response $f_{j^*} = \theta = 1.4$ as more subject-specific data are collected.

To illustrate how predictions in this model depend on the amount of subject-specific data collected, we conducted a simulation of model (4) for a particular individual. For this example, the parameter θ for the individual was chosen far from the population mean (when compared to the magnitude of the population variance), in order to make the transition from the population mean to the true expected response f_{j^*} large enough so as not to be obscured by measurement noise in the example. We assumed the population parameters $\mu = 0$ and $\delta = 1$ and chose the subject-specific parameter value $\theta = 1.4$. We simulated $m = 10$ data points for the individual, with a standard deviation of measurement error of $\sigma = 1$. The MMSE prediction \hat{f}_{j^*} was calculated by incorporating observations y_j iteratively. Figure 2 shows \hat{f}_{j^*} plotted against the number of data points used. The variance of the posterior distribution for f_{j^*} from (11) was used to construct confidence intervals on \hat{f}_{j^*} . As expected, for the individual considered, the prediction \hat{f}_{j^*} began at the population mean and moved to the true expected response with shrinking confidence interval as more simulated data were collected.

4. Bivariate Subject-Specific Bayesian Models

When subject-specific data are sparse, individualized predictions may not, on average, reach acceptable levels of accuracy. Accuracy may be improved by including data from a secondary subject-specific data source. However,

individual differences in one response variable may not be identical to individual differences in another (e.g., [6, 21]). Therefore, data from a secondary response variable may not simply be used as a substitute for the primary response variable. Rather, to improve prediction accuracy on the primary response by incorporating data from the secondary response, the between-subjects correlation between the primary and secondary response variables must be taken into account.

Here we derive the average accuracy of individualized predictions of a primary response variable based on distinct primary and secondary subject-specific response variables, accounting for between-subjects correlation between the two responses. Let i be an index for individual, let r be an index for response type, and let j be an index for observations ordered by time, nested within individual and response type. Suppose that the response y_{rij} can be modeled by

$$y_{rij} = f(t_{rij}, \theta_{ri}) + \epsilon_{rij}, \quad (15)$$

where $f(t_{rij}, \theta_{ri})$ represents the model function, t_{rij} represents measurement time, θ_{ri} represents a random subject-specific parameter associated with response type r , and ϵ_{rij} represents additive measurement error. Limiting our focus to a particular individual, we may remove the subscript i and model the subject's responses as

$$y_{rj} = f_{rj} + \epsilon_{rj}, \quad (16)$$

where f_{rj} is used to denote $f(t_{rj}, \theta_r)$. Suppose that a total of m_r responses have been observed for each response type r for the individual at hand. In the next sections, we focus on constructing the MMSE prediction for the expected primary response, \hat{f}_{1j^*} .

5. Bivariate Random Intercept Model

We consider the bivariate random intercept model obtained from (16) by letting $f_{rj} = \theta_r$:

$$y_{rj} = \theta_r + \epsilon_{rj}, \quad (17)$$

where $r = 1, 2$. The scalar model can be converted to vector form by concatenating the responses for each response type,

$$\mathbf{y}_1 = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1m_1} \end{pmatrix}, \quad (18)$$

$$\mathbf{y}_2 = \begin{pmatrix} y_{21} \\ \vdots \\ y_{2m_2} \end{pmatrix},$$

and then concatenating the response vectors of different types,

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}. \quad (19)$$

A similar assembly of the measurement errors can be done so that

$$\begin{aligned}\boldsymbol{\epsilon}_1 &= \begin{pmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1m_1} \end{pmatrix}, \\ \boldsymbol{\epsilon}_2 &= \begin{pmatrix} \epsilon_{21} \\ \vdots \\ \epsilon_{2m_2} \end{pmatrix}, \\ \boldsymbol{\epsilon} &= \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \end{pmatrix}.\end{aligned}\quad (20)$$

An assembly of the parameters can be accomplished by first constructing the parameter vector,

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \quad (21)$$

and the design matrix,

$$\mathbf{H} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}, \quad (22)$$

so that the single individual model of (17) can be vectorized as

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \boldsymbol{\epsilon}. \quad (23)$$

We consider the case where subject-specific traits and measurement errors are both normally distributed,

$$\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_\theta), \quad (24)$$

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\epsilon), \quad (25)$$

where $\boldsymbol{\mu}$, \mathbf{C}_θ , and \mathbf{C}_ϵ are fixed population characteristics.

Correlations between primary and secondary response variables \mathbf{y}_1 and \mathbf{y}_2 can be modeled as arising from a correlation between θ_1 and θ_2 (between-subjects correlation) or correlations between $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ (within-subjects correlation). Here, we assume that response correlations arise from between-subjects correlations:

$$\mathbf{C}_\theta = \begin{pmatrix} \delta_1^2 & \rho\delta_1\delta_2 \\ \rho\delta_1\delta_2 & \delta_2^2 \end{pmatrix}, \quad (26)$$

where ρ ($-1 < \rho < 1$) represents the between-subjects correlation between primary and secondary response variable means and δ_r^2 represents the between-subjects variance for response variable r . Furthermore, we assume that measurement errors are uncorrelated with response variable-specific variance σ_r^2 , so that correlations between the response variables arise only from the between-subjects components. For subject-specific repeated-measures data with no covariates for two response variables, it may be fair to consider the error variance within subjects to be independent as long as perturbations from the intercepts do not tend to be common over both response types.

The error covariance matrix for response type r is a diagonal matrix with dimension m_r , where each nonzero element is the type-specific error variance σ_r^2 ,

$$\boldsymbol{\Sigma}_r = \begin{pmatrix} \sigma_r^2 & 0 & \dots \\ 0 & \sigma_r^2 & \ddots \\ \vdots & \ddots & \ddots \end{pmatrix}. \quad (27)$$

The full error covariance matrix can then be built from the type-specific blocks,

$$\mathbf{C}_\epsilon = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{pmatrix}. \quad (28)$$

The bivariate Bayesian model we consider here is fully characterized by (23)–(28).

As was the case for model (4), for a particular individual, the expected primary response at time t_{1j^*} is

$$E[y_{1j^*} | \boldsymbol{\theta}] = f_{1j^*} = \theta_1. \quad (29)$$

The MMSE prediction is

$$\hat{f}_{1j^*} = \hat{\theta}_1. \quad (30)$$

The MMSE estimator for model (23) is [23]:

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\mu} + (\mathbf{C}_\theta^{-1} + \mathbf{H}'\mathbf{C}_\epsilon^{-1}\mathbf{H})^{-1} \mathbf{H}'\mathbf{C}_\epsilon^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu}). \quad (31)$$

The MMSE estimator $\hat{\theta}_1$ can be extracted as the first element of $\hat{\boldsymbol{\theta}}$.

The variance of the posterior distribution of $\boldsymbol{\theta} | \mathbf{y}$ is given by

$$\text{Var}(\boldsymbol{\theta} | \mathbf{y}) = (\mathbf{C}_\theta^{-1} + \mathbf{H}'\mathbf{C}_\epsilon^{-1}\mathbf{H})^{-1}. \quad (32)$$

The variance of the posterior distribution of $\theta_1 | \mathbf{y}$ (and therefore $f_{1j^*} | \mathbf{y}$) can be obtained by extracting the first element from the diagonal of $\boldsymbol{\theta} | \mathbf{y}$. The BMSE of the MMSE estimator $\boldsymbol{\theta}$ can be obtained from the parameter BMSE matrix $\mathbf{M}_{\hat{\boldsymbol{\theta}}}$, which is given by [23]:

$$\mathbf{M}_{\hat{\boldsymbol{\theta}}} = \text{Var}(\boldsymbol{\theta} | \mathbf{y}). \quad (33)$$

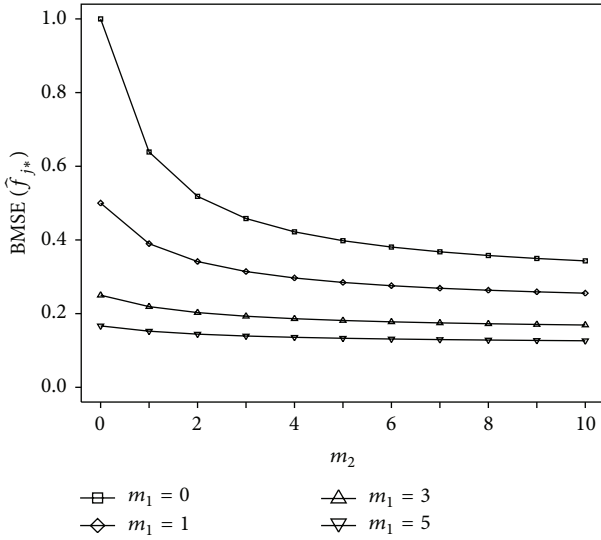


FIGURE 3: BMSE (i.e., accuracy) of MMSE predictions of the expected primary response \hat{f}_{1j^*} for the bivariate random intercept model specified in (23), as a function of the amount of data collected on a secondary response variable (m_2), shown for different amounts of data collected on the primary response variable (m_1). The figure illustrates that the accuracy of individualized Bayesian forecasting predictions improves progressively with just a few measurements of the secondary response variable when measurements of the primary response variable are increasingly sparse.

The BMSE of $\hat{\theta}_1$ (and therefore \hat{f}_{1j^*}) can be obtained by extracting the first element from the diagonal of $\mathbf{M}_{\hat{\theta}}$. Substituting (22), (26) and (28) into (33), we find that the BMSE for $\hat{\theta}_1$ can be simplified as follows:

$$\mathbf{M}_{\hat{\theta}_1} = \left(\frac{m_1}{\sigma_1^2} + \frac{1}{\delta_1^2} + \lambda(m_2) \right)^{-1}, \quad (34)$$

where

$$\lambda(m_2) = \frac{\rho^2 \delta_2^2 / \delta_1^2}{\delta_2^2 (1 - \rho^2) + \sigma_2^2 / m_2}. \quad (35)$$

Figure 3 illustrates the dependence of the BMSE of \hat{f}_{1j^*} on the number of observations from the primary and secondary responses. For this illustration, the population parameters were fixed at the values $\delta_1 = 1$, $\delta_2 = 1$, $\rho = 0.85$, $\sigma_1 = 1$, and $\sigma_2 = 1$. The figure shows the decrease of the BMSE as a function of the collection of secondary response measurements, given different numbers of primary response measurements. For a large number of primary response measurements, little change in BMSE is derived from the secondary response. However, for a small number of primary response measurements, the BMSE decreases substantially with just a few measurements of the secondary response.

To further illustrate the bivariate Bayesian forecasting procedure, simulated subject-specific parameter pairs $\boldsymbol{\theta} = (\theta_1, \theta_2)$ from the population distribution given by (24) with

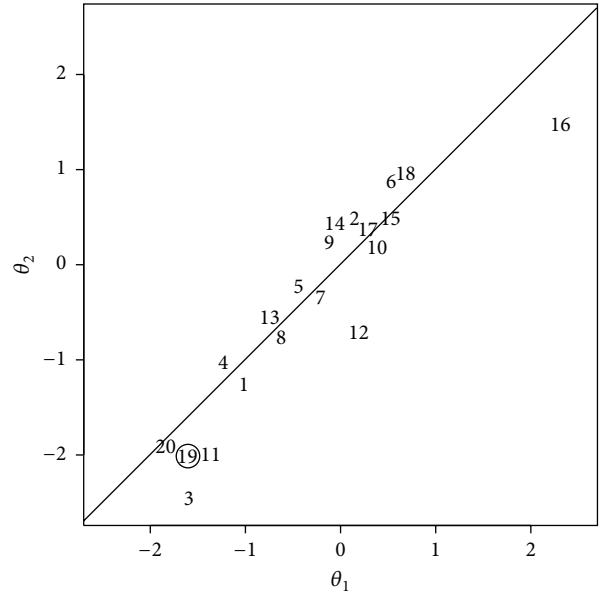


FIGURE 4: Subject-specific parameter pairs, simulated from a probability distribution specified by (24), from which a simulated individual was selected for a bivariate Bayesian forecasting simulation. Each number represents the parameter pair $\boldsymbol{\theta} = (\theta_1, \theta_2)$ for a different simulated individual. The circled individual (#19) is used for the illustration in Figure 5. The diagonal line shows where the points would fall given a between-subjects correlation of $\rho = 1$.

$\delta_1 = 1$, $\delta_2 = 1$, and $\rho = 0.9$ were generated for $N = 20$ individuals, as shown in Figure 4. From this simulated set, individual #19 (circled in Figure 4) with subject-specific parameter vector $\boldsymbol{\theta} = (-1.6; -2.0)$ was chosen to illustrate the transition of the primary response prediction from the population mean response to the true expected response f_{1j^*} through Bayesian forecasting. For this individual, we simulated errors from (25) using $\sigma_1 = \sqrt{2}$, $\sigma_2 = \sqrt{0.5}$, $m_1 = 10$, and $m_2 = 10$. Bivariate responses were then constructed using (23).

The MMSE prediction \hat{f}_{1j^*} for individual #19 was iteratively determined after assuming only primary responses were observed and after assuming pairs of primary and secondary responses were observed. The iterative estimates for both cases are shown in Figure 5, along with the simulated data. The variance of the posterior distribution for f_{1j^*} from (32) was used to construct 95% confidence intervals on f_{1j^*} . For the individual considered, the predictions based on both response variables (purple line) were more accurate than the predictions based on only the primary response (blue line) for the first few iterations of MMSE prediction. Note, however, that while this behavior of the prediction accuracy is found on average, as follows from (34) it is not necessarily true for each and every individual to which the procedure may be applied. Thus, caution is needed in relying on bivariate Bayesian forecasting for improved prediction accuracy of specific individuals; improved prediction accuracy can only be counted on in the average over individuals.

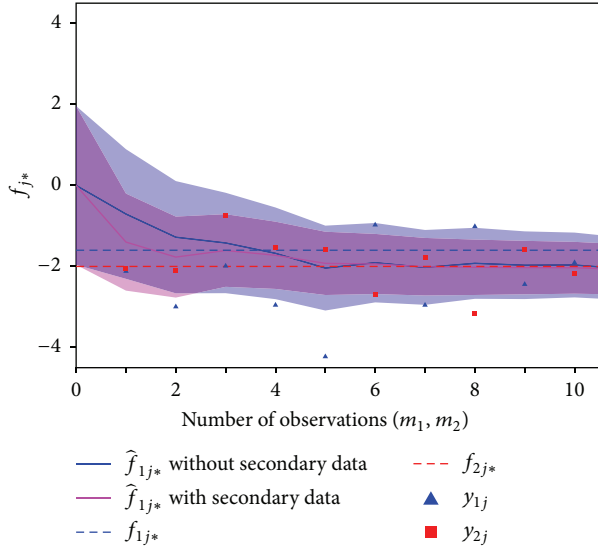


FIGURE 5: MMSE predictions of the expected primary response \hat{f}_{1j^*} from the bivariate random intercept model specified in (23), and corresponding 95% confidence intervals (shaded areas) for individual #19. The MMSE estimator \hat{f}_{1j^*} was iteratively determined assuming only primary responses were observed, as well as assuming both primary and secondary responses were observed. For the former case, the MMSE was iteratively determined by incorporating observations y_{1j} ; for the latter case, the MMSE was iteratively determined by incorporating pairs of observations (y_{1j}, y_{2j}) . Confidence intervals were obtained from quantiles of the posterior distribution, which is defined by the posterior mean (the MMSE estimator) (31) and the posterior variance (32).

6. Data Collection Cost Minimization

When Bayesian forecasting is applied to individualize predictions, data must be collected to tailor the population model to the individual at hand. In certain sleep research applications, such as forecasting of sleep parameters across nights or predicting performance deficits across periods of sleep deprivation, and in a wide range of other biomedical contexts, this requires creating multiple opportunities for taking measurements. This may be an expensive proposition, and reducing the number of measurement bouts needed to obtain the necessary data could entail considerable cost savings. By measuring secondary responses and incorporating these through bivariate Bayesian forecasting, it may be possible to achieve a given level of prediction accuracy at lower overall cost of data acquisition. Here we explore this possibility in the case of the bivariate random intercept model.

We consider a scenario in which the cost of collecting an observation on the primary response is c_1 , the cost of collecting an observation on the secondary response is c_2 , and the total cost of data collection is the sum of primary and secondary response costs accrued,

$$c = m_1 c_1 + m_2 c_2, \quad (36)$$

where $m_1, m_2 \geq 0$. For this scenario, we determine how many observations m_1 and m_2 we may expect to have to collect from each response type in order to minimize the total cost of achieving, on average, a given prediction accuracy η^2 :

$$\mathbf{M}_{\hat{f}_{1j^*}} = \eta^2. \quad (37)$$

We can simplify the minimization problem by removing either m_1 or m_2 from both the total cost equation and the nonnegativity constraints on the number of data points. Using $\hat{f}_{1j^*} = \hat{\theta}_1$ (see (30)), it follows that the BMSE of \hat{f}_{1j^*} is equal to the BMSE of $\hat{\theta}_1$ (i.e., $\mathbf{M}_{\hat{f}_{1j^*}} = \mathbf{M}_{\hat{\theta}_1}$). Fixing $\mathbf{M}_{\hat{f}_{1j^*}} = \eta^2$ (37) therefore implies that $\mathbf{M}_{\hat{\theta}_1} = \eta^2$, which can be used with (34) to obtain a relationship between the number of primary and secondary observations needed to meet the average accuracy criterion η^2 :

$$m_1 = \sigma_1^2 \left(\frac{1}{\eta^2} - \frac{1}{\delta_1^2} - \lambda(m_2) \right). \quad (38)$$

We then substitute (38) into (36):

$$c = c_1 \sigma_1^2 \left(\frac{1}{\eta^2} - \frac{1}{\delta_1^2} - \lambda(m_2) \right) + c_2 m_2. \quad (39)$$

The constraint $m_1 \geq 0$ can be equivalently formulated as an upper bound on $\lambda(m_2)$:

$$\lambda(m_2) \leq \frac{1}{\eta^2} - \frac{1}{\delta_1^2}. \quad (40)$$

Consideration of this constraint is only necessary if there are a certain number of measurements on the secondary response for which it is possible to obtain the desired accuracy without measurement of the primary response; that is,

$$\lim_{m_2 \rightarrow \infty} \lambda(m_2) \geq \frac{1}{\eta^2} - \frac{1}{\delta_1^2}. \quad (41)$$

Substituting for $\lambda(m_2)$ from (35), this latter condition can be reformulated as follows:

$$\delta_1^2 (1 - \rho^2) \leq \eta^2. \quad (42)$$

Thus, m_2 has an upper bound when η^2 , the desired BMSE of \hat{f}_{1j^*} , is not smaller than $\delta_1^2(1 - \rho^2)$, the minimum BMSE of \hat{f}_{1j^*} which can be obtained using only the secondary response. It follows that the constraint set for the minimization problem is

$$\begin{aligned} 0 \leq m_2 &\leq \frac{\sigma_2^2 (1/\eta^2 - 1/\delta_1^2)}{\delta_2^2 (1 - \rho^2) (1/\delta_1^2 (1 - \rho^2) - 1/\eta^2)} \quad \text{if } \delta_1^2 (1 - \rho^2) \leq \eta^2, \\ 0 \leq m_2 &\text{ otherwise.} \end{aligned} \quad (43)$$

If m_2 exceeds its upper bound then the number of secondary response measurements is more than what is minimally needed to meet the average accuracy criterion (37).

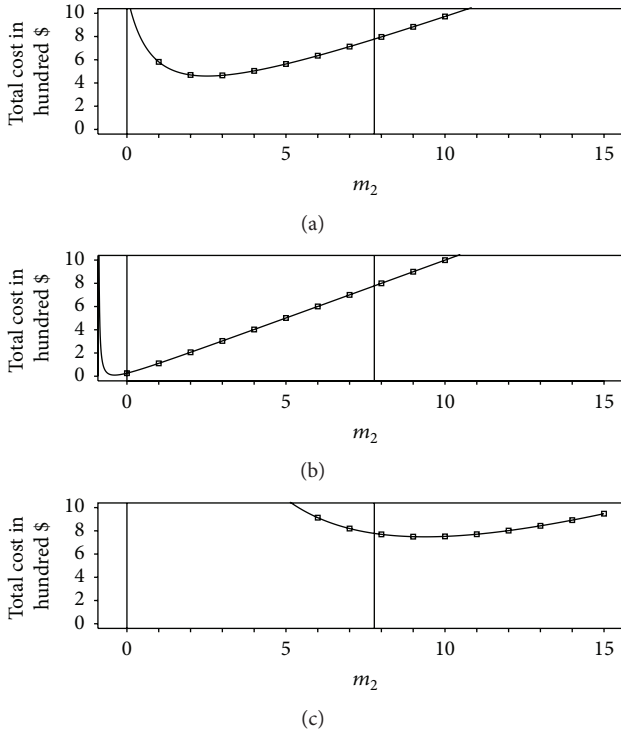


FIGURE 6: Illustration of three types of absolute minima of the total cost of obtaining a fixed average prediction accuracy inside the region defined by (43). Each plot shows the total cost $c = c_1 m_1 + c_2 m_2$ plotted against secondary response sample size m_2 . For m_2 given, the primary response sample size m_1 that obtains a fixed BMSE of η^2 in the expected primary response \hat{f}_{j^*} is computed using (38). The lower and upper boundaries of the region defined by (43) are shown with solid vertical lines. (a) shows an interior point minimum, obtained by letting $\sigma_1 = 1.00$. This type of solution occurs when the minimum of the unconstrained cost function lies within the region defined by (43). (b) shows a lower boundary solution, obtained by letting $\sigma_1 = 0.15$. This type of solution occurs when the minimum of the unconstrained cost function lies below the feasible region defined by (43). (c) shows an upper boundary solution, obtained by letting $\sigma_1 = 3.00$. This type of solution occurs when the minimum of the unconstrained cost function lies above the region defined by (43).

The minimal cost solution occurs either on the boundary of the region defined by (43) or at a local minimum in the interior of this region. Figure 6 shows how different values of the error variance σ^2 can result in either boundary or interior solution types. For this demonstration, the population parameters are set at $\rho = 0.85$, $\delta_1 = 1.0$, $\delta_2 = 1.0$, and $\sigma_2 = 0.5$, the costs of measurement are assumed to be $c_1 = \$500$ and $c_2 = \$100$, and the BMSE of \hat{f}_{j^*} is fixed at the value $\eta^2 = 0.30$.

In cases where the solution lies in the interior of (43) at a local minimum, the solution must occur at critical points of (39), which can be found by setting to zero the derivative of the total cost with respect to m_2 :

$$\frac{\partial}{\partial m_2} c(m_2) = -c_1 \sigma_1^2 \frac{\partial \lambda(m_2)}{\partial m_2} + c_2 = 0, \quad (44)$$

where

$$\frac{\partial \lambda(m_2)}{\partial m_2} = \frac{\sigma_2^2 \rho^2 \delta_2^2 / \delta_1^2}{(m_2 \delta_2^2 (1 - \rho^2) + \sigma_2^2)^2}. \quad (45)$$

Substituting the above expression for $\partial \lambda(m_2) / \partial m_2$ into (44) and solving for m_2 , we obtain the following two critical points:

$$m_2^\pm = \frac{\sigma_2}{\delta_1 \delta_2^2 (1 - \rho^2)} \left(-\delta_1 \sigma_2 \pm \sqrt{\frac{c_1}{c_2} \delta_2 \sigma_1 |\rho|} \right). \quad (46)$$

The smaller critical point m_2^- can be disregarded as a possible solution since it is always less than zero. The second derivative at m_2^+ ,

$$\frac{\partial^2}{\partial m_2^2} c(m_2^+) = \frac{2c_2 \sqrt{c_2/c_1} (1 - \rho^2) \delta_1 \delta_2}{|\rho| \sigma_1 \sigma_2}, \quad (47)$$

is positive when $|\rho| < 1$, which implies that the cost function exhibits a local minimum at this point. If the local minimum m_2^+ is inside the region defined by (43) (see Figure 6(a)), then the solution to the cost minimization problem is

$$\hat{m}_1 = \sigma_1^2 \left(\frac{1}{\eta^2} - \frac{1}{\delta_1^2 (1 - \rho^2)} \left(1 - \sqrt{\frac{c_2 \delta_1^2 / \sigma_1^2}{c_1 \delta_2^2 / \sigma_2^2} \rho^2} \right) \right), \quad (48)$$

$$\hat{m}_2 = m_2^+,$$

where \hat{m}_1 is determined by substituting \hat{m}_2 into (38), and the total cost is found from (36).

Alternatively, if m_2^+ is below the lower boundary of the region defined by (43) (see Figure 6(b)), then the minimal cost solution involves collecting no data from the secondary response. The conditions for which the secondary response is not part of the minimal cost solution are as follows:

$$\frac{c_2}{c_1} > \frac{\delta_2^2 / \sigma_2^2}{\delta_1^2 / \sigma_1^2} \rho^2, \quad (49)$$

where δ_r^2 / σ_r^2 reflects the between-to-within variance ratio for the r th response type. For this case, the solution which achieves, on average, the level of accuracy η^2 is found from (38):

$$\hat{m}_1 = \sigma_1^2 \left(\frac{1}{\eta^2} - \frac{1}{\delta_1^2} \right), \quad (50)$$

$$\hat{m}_2 = 0.$$

Finally, if m_2^+ is above the upper boundary of the region defined by (43) (see Figure 6(c)), then the solution for m_2 occurs at this boundary, where all the data are collected from the secondary response and none from the primary response. For this case, the solution is as follows:

$$\hat{m}_1 = 0,$$

$$\hat{m}_2 = \frac{\sigma_2^2 (1/\eta^2 - 1/\delta_1^2)}{\delta_2^2 (1 - \rho^2) (1/\delta_1^2 (1 - \rho^2) - 1/\eta^2)}. \quad (51)$$

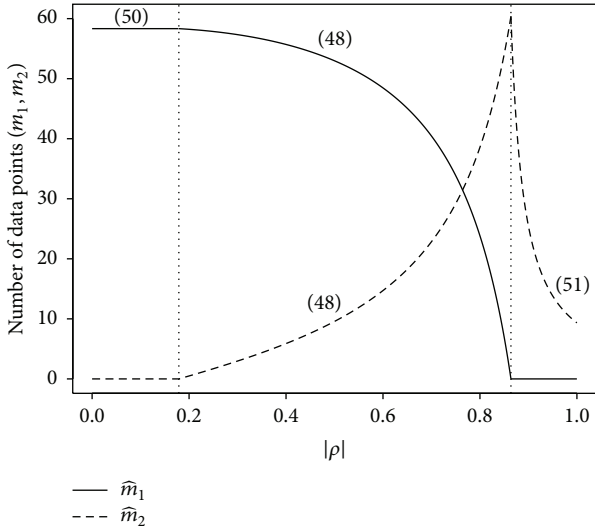


FIGURE 7: Example of data collection cost minimization that shows the number of observations on primary and secondary responses needed to obtain a fixed BMSE in the MMSE prediction of the expected primary response \hat{f}_{j^*} , for different values of the between-subjects correlation between primary and secondary response parameters. The solid curve represents the number of measurements to collect from the primary response, \hat{m}_1 , and the dashed curve represents the number of measurements to collect from the secondary response, \hat{m}_2 . In this example, for $0.0 \leq |\rho| \leq 0.18$, no data is to be collected from the secondary response, and the number of data points to collect from the primary response is obtained from (50). For $0.18 < |\rho| \leq 0.85$, the number of secondary observations increases and the number of primary observations decreases with increasing correlation, as specified by \hat{m}_2 and \hat{m}_1 in (48). For $0.85 < |\rho| \leq 1.0$, observations are to be collected only from the secondary response, the number of which is given by (51). The equation numbers are indicated near the relevant pieces of the curves.

Figure 7 illustrates the number of observations required from primary and secondary responses to obtain a fixed level of accuracy η^2 on average for different values of the between-subjects correlation ρ between primary and secondary response parameters. For the example shown, the population model parameters and cost parameters were fixed at $\delta_1 = 1$, $\delta_2 = 1$, $\sigma_1 = 1$, $\sigma_2 = 0.5$, $c_1 = 5$, and $c_2 = 1$, and the desired level of accuracy was $\eta^2 = 0.30$. The figure illustrates the three cases described by (48), (50), and (51).

7. Example: Efficient Assessment of an Individual's Characteristic Wakefulness after Sleep Onset

To illustrate the cost minimization approach outlined in the previous section, we apply it in an example involving the assessment of wakefulness after sleep onset (WASO) in laboratory-based sleep studies. Here we define WASO as the duration of intermittent wakefulness during a sleep period,

between the time of sleep onset and the time of final awakening. WASO can be measured by polysomnography (PSG), that is, measuring the sleep electroencephalogram (EEG) and other physiological sleep signals and scoring sleep/wake states, typically in 30 s epochs, based on those signals. PSG is the gold standard procedure for sleep/wake assessment, but it is labor-intensive and expensive to perform. WASO may also be measured in the laboratory using wrist actigraphy (i.e., wrist activity monitoring), which is considerably less expensive. Although actigraphy is not considered a gold standard for measuring WASO, the correspondence with PSG-based WASO is at least moderate in healthy populations [24].

We base our example on data from $n = 33$ subjects (ages 22–38; 15 females) who spent between 6 and 13 nights and days inside a controlled laboratory environment with 10 h in bed for sleep (22:00–08:00) each day. The Institutional Review Board (IRB) of Washington State University approved the research, and subjects gave written informed consent. WASO was measured using both PSG (WASO-P) and actigraphy (WASO-A). PSG recordings were performed using digital equipment (Nihon Kohden, Foothill Ranch, CA). Sleep stages and periods of wakefulness were scored in 30 s epochs using standard criteria [25], and WASO-P was calculated from the scored records. Actigraphic recordings were made with Motionlogger wrist actigraphs (Ambulatory Monitoring, Inc., Ardsley, NY). Sleep and wakefulness were assessed from the actigraphic records using the automated algorithm of [26], which calculated WASO-A.

Let WASO-P be the primary response (as it is the gold standard measure) and let WASO-A be the secondary response. Assuming that WASO-P and WASO-A are normally distributed around distinct subject-specific means, we apply the model defined by (23)–(28) to our example. We anticipate that the subject-specific means for WASO-P and WASO-A are positively correlated. We aim to determine a cost-effective data collection scheme given a specific level of desired accuracy in estimates of an individual's mean WASO-P. For illustration purposes, we assume a fixed cost of \$1,250 per night for WASO-P and \$150 per night for WASO-A. We wish to estimate an individual's mean WASO-P to an average accuracy (i.e., $\sqrt{\text{BMSE}}$) of $\eta = 15$ min. Using the equations derived in the previous section, we estimate the number of nights of WASO-P and WASO-A that most cost-effectively achieves the desired level of accuracy on average.

We have estimated the population model parameters using the nlme package for R [27]. This package fits mixed-effects models using the approach of [28]. The maximum likelihood estimation method tends to underestimate variance parameters [29]; therefore, we estimate these parameters using the restricted maximum likelihood method [6].

In our dataset, the overall means for WASO-P and WASO-A were estimated as follows: $\mu_1 = 54 \text{ min} \pm 5 \text{ min}$ and $\mu_2 = 32 \text{ min} \pm 4 \text{ min}$ (estimate \pm standard error), indicating that WASO-A tended to underestimate the total amount of WASO as compared to PSG. The estimated variability between subjects for WASO-P and WASO-A was found to be the same: $\delta_1 = \delta_2 = 21 \text{ min}$. There was a substantial

correlation between subject means for WASO-P and WASO-A: $\rho = 0.69$ (95% confidence interval: [0.30, 0.90]). The within-subject variation around the subject mean was $\sigma_1 = 32$ min for WASO-P and $\sigma_2 = 21$ min for WASO-A.

We determine from (43) that m_2 , the number of actigraphy nights, is not bounded above; that is, we cannot achieve our accuracy with actigraphy alone. Further, we find from (46) that m_2^+ is within the feasible region defined by (43), and, therefore, the solution to the cost minimization problem is given by (48). Applying these equations, we achieve an average accuracy of $\eta = 15$ min for minimal cost by collecting 3.99 nights of actigraphy and 0.73 nights of PSG (see Figure 8(a)). An approximately optimal solution in the integer domain is found by the common practice of rounding the optimal continuous solution to the nearest integer values [30]. We verify through a grid search that the minimal cost integer value solution neighbors the analytic solution and can be obtained by rounding down to three nights of actigraphy and up to one night of PSG. This yields a total cost of \$1700. In contrast, to achieve the same accuracy with PSG alone would require 2.24, or in practice three nights of measurement, for a total cost of \$3750.

Note that the results are highly dependent on the estimated between-subjects correlation and that the 95% confidence interval for this correlation was large. The minimal cost solution also depends on the level of accuracy that is required; see Figures 8(b) and 8(c) for scenarios with an average accuracy of 14 min and 13 min, respectively.

8. Linear Models with Time Dependency

For both the univariate and bivariate linear models, time dependency can be introduced by adding time as a covariate. This complicates the construction of a design matrix that enables predictions with a given average accuracy. We show that, in models with time dependency, the BMSE of a predicted response depends on the times at which responses are measured, and this dependency can be summarized by the mean and variance of the measurement times.

To illustrate, we consider a linear approximation of a time-dependent model known as the two-process model of sleep regulation [10, 31]. It has been shown that, for a range of sleep/wake scenarios, the two-process model can describe temporal changes in waking cognitive performance as the algebraic difference between two functions describing physiological processes: the homeostatic pressure for sleep and the circadian pressure for wakefulness [11]. Here we focus solely on modeling the homeostatic pressure for sleep, the dynamics of which are specified separately for sleep and for wakefulness. The dynamics can be modeled using the recursive formulation of [10]

$$S_t = \begin{cases} e^{-\Delta t/\tau_d} S_{t-1} & \text{(sleep)} \\ 1 - e^{-\Delta t/\tau_r} (1 - S_{t-1}) & \text{(wake)}, \end{cases} \quad (52)$$

where S_t represents the homeostatic pressure after the t th time step of duration Δt , S_{t-1} represents the homeostatic pressure at one time step before (i.e., at time $(t-1) \cdot \Delta t$), Δt

is typically fixed at 0.5 h, and τ_d and τ_r are time constants for the decay and rise of the homeostatic process during sleep and wakefulness, respectively.

We divide the sleep/wake schedule into periods of sleep and periods of wake, both indexed by k . Let $S_0^{(k)}$ represent the initial homeostatic pressure for the k th period of wakefulness. It has been proposed that change over time in the model is better modeled as linear rather than exponential [32]. For the purpose of the present example, we adopt this idea and approximate the model over a period of continuous wakefulness using a linear interpolation between the start and end points of the wake period (see Figure 9).

Let y_j represent (hypothetical) measurements of the build-up of homeostatic pressure for sleep during wakefulness. These data can be modeled using the following approximation:

$$y_j = S_0^{(k)} + \frac{(1 - e^{-T^{(k)}/\tau_r})(1 - S_0^{(k)})}{T^{(k)}} t_j + \epsilon_j, \quad (53)$$

where t_j represents the amount of time elapsed since awakening, $T^{(k)}$ is used to denote the duration of the k th wake period, and ϵ_j represents normally distributed measurement error.

We consider the homeostatic process over a repeating schedule consisting of $T^{(k)} = T = 16$ h of wakefulness and 8 h of sleep. In this schedule, individuals maintain a steady state for which the homeostatic pressure $S_0^{(k)}$ at the onset of the wake period is constant across days; that is, $S_0^{(k)} = S_0$. This allows us to derive the homeostatic pressure at the start of wakefulness as a function of τ_r and τ_d :

$$S_0 = \frac{e^{-(24-T)/\tau_d} (1 - e^{-T/\tau_r})}{1 - e^{-((24-T)/\tau_d + T/\tau_r)}}. \quad (54)$$

The equation for the homeostatic pressure during a particular wake period can thus be written as

$$y_j = \alpha + \beta t_j + \epsilon, \quad (55)$$

where

$$\alpha = S_0, \quad (56)$$

$$\beta = \frac{(1 - e^{-T/\tau_r})(1 - S_0)}{T}. \quad (57)$$

In matrix form, the model can be written as

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad (58)$$

where the design matrix is given by

$$\mathbf{H} = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix}, \quad (59)$$

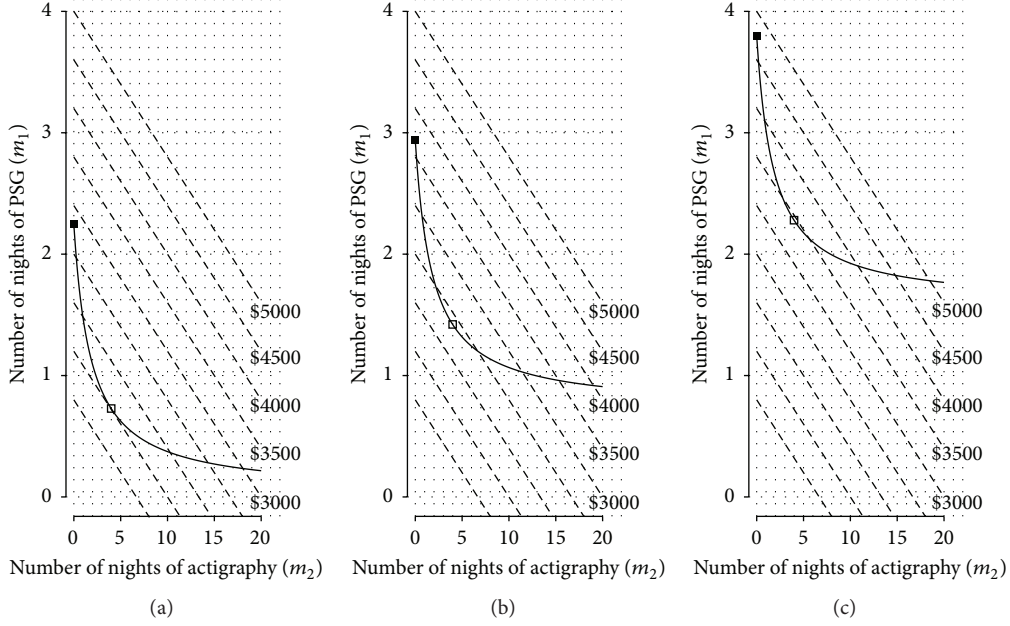


FIGURE 8: Cost of collection and number of nights of actigraphy and polysomnography (PSG) which combined will yield a certain average accuracy of wakefulness after sleep onset (WASO) parameter estimates. For illustration purposes, PSG and actigraphy are assumed to cost \$1250 and \$150 per night, respectively. The cost of data collection is fixed on each diagonal dashed line (for illustrative purposes these are only shown at fixed intervals) and increases as we collect more nights of PSG and actigraphy. Subject-specific WASO estimates of an average accuracy of (a) 15 min, (b) 14 min, and (c) 13 min are obtained on the solid curve. The point that minimizes the cost of obtaining the fixed accuracy (open squares) is determined as the point on the fixed accuracy curve where the line tangent to the curve is parallel to the fixed cost lines. For comparison the solution for obtaining the same average accuracy using only polysomnography is also shown (solid squares).

and the parameter vector is given by

$$\boldsymbol{\theta} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (60)$$

Analogous to (5), but for two parameters, we assume the following prior distribution on model parameters:

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \sim \mathcal{N} \left(\boldsymbol{\mu}, \begin{pmatrix} \delta_\alpha^2 & 0 \\ 0 & \delta_\beta^2 \end{pmatrix} \right). \quad (61)$$

As in (6), we assume that the errors are independent realizations from a normal distribution with zero mean and variance σ^2 . Analogous to (11) and (12), the BMSE for the MMSE prediction of the expected (primary) response \hat{f}_{j^*} at some time t_{j^*} in the univariate case is as follows (see Appendix A):

$$\begin{aligned} \mathbf{M}_{\hat{f}_{j^*}} &= \frac{1/\delta_\beta^2 + ms^2/\sigma^2 + m(\bar{t} - t_{j^*})^2/\sigma^2 + t_{j^*}^2/\delta_\alpha^2}{(1/\delta_\alpha^2 + m/\sigma^2)(1/\delta_\beta^2 + m(s^2 + \bar{t}^2)/\sigma^2) - (m\bar{t}/\sigma^2)^2}, \end{aligned} \quad (62)$$

where \bar{t} denotes the mean of these times,

$$\bar{t} = \frac{1}{m} \sum_{j=1}^m t_j, \quad (63)$$

and s^2 denotes the variance of these times,

$$s^2 = \frac{1}{m} \sum_{j=1}^m (t_j - \bar{t})^2. \quad (64)$$

Our task is to determine the measurement times t_1, \dots, t_m for this example that will minimize $\mathbf{M}_{\hat{f}_{j^*}}$. Equation (62) shows that $\mathbf{M}_{\hat{f}_{j^*}}$ depends on the measurement times only through their mean \bar{t} and variance s^2 . Consequently, instead of conducting an m -dimensional minimization of $\mathbf{M}_{\hat{f}_{j^*}}$ over all the measurement times t_1, \dots, t_m , we can write $\mathbf{M}_{\hat{f}_{j^*}}$ as a function of \bar{t} and s^2 and conduct a two-dimensional minimization. In doing this, we find that $\mathbf{M}_{\hat{f}_{j^*}}$ is minimized both when $s^2 \rightarrow \infty$ (see Appendix B) and also, more practically relevant, when

$$\bar{t} = \frac{t_{j^*} (\sigma^2/m + \delta_\alpha^2)}{\delta_\alpha^2}. \quad (65)$$

The optimal mean measurement time for this example, given by (65), lies slightly above the prediction time; in the limit as $\delta_\alpha^2 \rightarrow \infty$, $\mathbf{M}_{\hat{f}_{j^*}}$ is minimized when the data are collected at times such that $\bar{t}_{\min} = t_{j^*}$. If the prior variance on the intercept is equal to the error variance (i.e.,

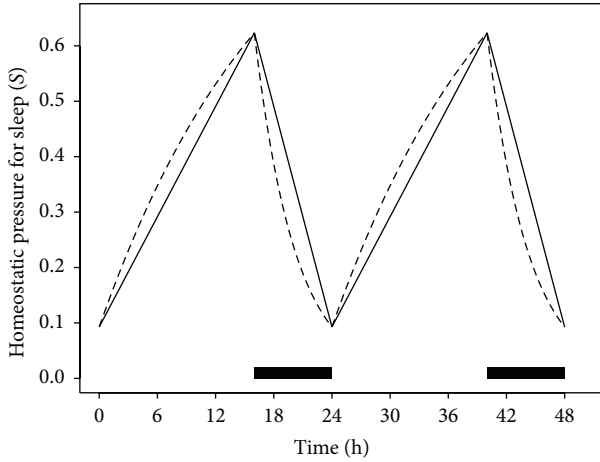


FIGURE 9: Homeostatic pressure for wakefulness plotted over two complete sleep/wake cycles for a repeating schedule with 16 h of wakefulness and 8 h time in bed for sleep in each cycle. The dashed line represents the homeostatic pressure for sleep as given by (52); the solid line is a linear approximation based on interpolation between the sleep/wake transition points. The parameter values $\tau_d = 4.2$ h and $\tau_r = 18.2$ h are taken from [10]. The initial condition (S_0) is derived assuming steady state (54). Black bars indicate the 8 h periods in bed for sleep.

$\delta_\alpha^2 = \sigma^2$), the effect of the Bayesian prior is equivalent to increasing the value of \bar{t} that produces the minimum $\mathbf{M}_{\hat{f}_{j^*}}$ by t_{j^*}/m . This adjustment is the same as what would manifest with no prior information on the intercept when adding an additional measurement time, $t_{m+1} = 0$, to the design matrix \mathbf{H} .

The absolute minimum of $\mathbf{M}_{\hat{f}_{j^*}}$ in (62) is not always located in the feasible region in this example, as defined by $0 \leq t_j \leq T, \forall j \in \{1, \dots, m\}$. More specifically, the absolute minimum of the unconstrained case is not located inside the feasible region if and only if (see Appendix B):

$$\bar{t}_{\min} \geq T. \quad (66)$$

Under this condition we hypothesize that, within the feasible region, $\mathbf{M}_{\hat{f}_{j^*}}$ exhibits an absolute minimum when all the data are collected at time T . This is easy to show for the case of one measurement time (i.e., $m = 1$), and Appendix C contains a proof for the case of two measurement times (i.e., $m = 2$).

For $m > 2$, we conducted a simulation study to search for a counterexample (i.e., a case where the value of $\mathbf{M}_{\hat{f}_{j^*}}$ when all data is collected at time T is not the smallest value of $\mathbf{M}_{\hat{f}_{j^*}}$ within the feasible region). For the simulation study, we simulated 10,000,000 times with the following values:

$$\delta_\alpha \sim \text{Uniform}(0.001, 1);$$

$$\delta_\beta \sim \text{Uniform}(0.001, 1);$$

$$\sigma \sim \text{Uniform}(0.001, 1);$$

$$m \sim \text{Discrete Uniform}(1, 2, \dots, 100);$$

$$t_{j^*} \sim \text{Uniform}\left(T \frac{\delta_\alpha^2}{\sigma^2/m + \delta_\alpha^2}, T\right). \quad (67)$$

Concerning the ranges of the variance components, note that $\mathbf{M}_{\hat{f}_{j^*}}$ is invariant to the scale of the response. This can be demonstrated by scaling the variance matrices \mathbf{C}_θ and \mathbf{C}_ϵ by c_y and showing that the resulting $\mathbf{M}_{\hat{f}_{j^*}}$ is then scaled by the same factor c_y . The conclusion is that the shape of the surface of $\mathbf{M}_{\hat{f}_{j^*}}$ depends on the variance components only through their relative magnitudes. Furthermore, we argue that if any variance component is more than three orders of magnitude greater than any other component, then it would be advantageous to simplify the model by removing the smaller component. As such, all cases for which this model is reasonable can be covered within the range from 0.001 to 1 for each variance component. Further, concerning the number of observations, we expect that if there is a counterexample, it can be found somewhere in the range $1 \leq m \leq 100$. Finally, the range for t_{j^*} is determined specifically so that the absolute minimum lies outside the feasible region.

For each simulation, we compared $\mathbf{M}_{\hat{f}_{j^*}}$ at the hypothesized minimum, where $t_j = T, \forall j \in \{1, \dots, m\}$, to a randomly chosen time point, where each t_j is a realization from the following distribution:

$$t_j \sim \text{Uniform}(0, T). \quad (68)$$

For each of the 10,000,000 simulations, $\mathbf{M}_{\hat{f}_{j^*}}$ at the hypothesized minimum was indeed smaller than $\mathbf{M}_{\hat{f}_{j^*}}$ at the randomly chosen time point. Thus, we found no evidence against our original hypothesis that $\mathbf{M}_{\hat{f}_{j^*}}$ exhibits an absolute minimum if all the data are collected at time T . An analytical proof is beyond the scope of this paper.

We now extend our analysis to consider a time-dependent model with both primary and secondary responses. We formulate the model as follows:

$$y_{rj} = s_{rj} + \epsilon_{rj} = \alpha_r + \beta_r t_{rj} + \epsilon_{rj}. \quad (69)$$

In matrix form, the model is

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad (70)$$

where the design matrix is given by

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & t_{11} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & t_{1m_1} & 0 \\ 0 & 1 & 0 & t_{21} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & t_{2m_2} \end{pmatrix}, \quad (71)$$

and the parameter vector is given by

$$\boldsymbol{\theta} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{pmatrix}. \quad (72)$$

Let us assume the following prior distribution on model parameters:

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{pmatrix} \sim \mathcal{N} \left(\boldsymbol{\mu}, \begin{pmatrix} \delta_{\alpha_1}^2 & \rho\delta_{\alpha_1}\delta_{\alpha_2} & 0 & 0 \\ \rho\delta_{\alpha_1}\delta_{\alpha_2} & \delta_{\alpha_2}^2 & 0 & 0 \\ 0 & 0 & \delta_{\beta_1}^2 & \omega\delta_{\beta_1}\delta_{\beta_2} \\ 0 & 0 & \omega\delta_{\beta_1}\delta_{\beta_2} & \delta_{\beta_2}^2 \end{pmatrix} \right). \quad (73)$$

As in (28), we assume that the errors are independent realizations from a normal distribution with zero mean and σ_r ($r = 1, 2$). The BMSE of estimates of the primary response at time t_{1j^*} is given by

$$\mathbf{M}_{\hat{f}_{1j^*}} = \mathbf{h}' \mathbf{M}_{\hat{\boldsymbol{\theta}}} \mathbf{h}, \quad (74)$$

where

$$\mathbf{h} = \begin{pmatrix} 1 \\ 0 \\ t_{1j^*} \\ 0 \end{pmatrix}, \quad (75)$$

$$\mathbf{M}_{\hat{\boldsymbol{\theta}}} = \begin{pmatrix} \frac{1}{\delta_{\alpha_1}^2(1-\rho^2)} + \frac{m_1}{\sigma_1^2} & \frac{-\rho}{\delta_{\alpha_1}\delta_{\alpha_2}(1-\rho^2)} & \frac{m_1\bar{t}_1}{\sigma_1^2} & 0 \\ \frac{-\rho}{\delta_{\alpha_1}\delta_{\alpha_2}(1-\rho^2)} & \frac{1}{\alpha_2^2(1-\rho^2)} + \frac{m_2}{\sigma_2^2} & 0 & \frac{m_2\bar{t}_2}{\sigma_2^2} \\ \frac{m_1\bar{t}_1}{\sigma_1^2} & 0 & \frac{1}{\delta_{\beta_1}^2(1-\omega^2)} + \frac{m_1(s_1^2 + \bar{t}_1^2)}{\sigma_1^2} & \frac{-\omega}{\delta_{\beta_1}\delta_{\beta_2}(1-\omega^2)} \\ 0 & \frac{m_2\bar{t}_2}{\sigma_2^2} & \frac{-\omega}{\delta_{\beta_1}\delta_{\beta_2}(1-\omega^2)} & \frac{1}{\delta_{\beta_2}^2(1-\omega^2)} + \frac{m_2(s_2^2 + \bar{t}_2^2)}{\sigma_2^2} \end{pmatrix}^{-1}, \quad (76)$$

where \bar{t}_1 denotes the mean of the primary measurement times,

$$\bar{t}_1 = \frac{1}{m_1} \sum_{j=1}^{m_1} t_{1j}, \quad (77)$$

\bar{t}_2 denotes the mean of the secondary measurement times,

$$\bar{t}_2 = \frac{1}{m_2} \sum_{j=1}^{m_2} t_{2j}, \quad (78)$$

s_1^2 denotes the variance of the primary measurement times,

$$s_1^2 = \frac{1}{m_1} \sum_{j=1}^{m_1} (t_{1j} - \bar{t}_1)^2, \quad (79)$$

and s_2^2 denotes the variance of the secondary measurement times,

$$s_2^2 = \frac{1}{m_2} \sum_{j=1}^{m_2} (t_{2j} - \bar{t}_2)^2. \quad (80)$$

Our task in the bivariate case of the example is to determine the primary measurement times t_{11}, \dots, t_{1m_1} and

the secondary measurement times t_{21}, \dots, t_{2m_2} that minimize $\mathbf{M}_{\hat{f}_{1j^*}}$. Equations (74) and (76) show that $\mathbf{M}_{\hat{f}_{1j^*}}$ depends on the measurement times only through their response-specific means and variances, $\bar{t}_1, \bar{t}_2, s_1^2,$ and s_2^2 . Consequently, it is sufficient to minimize $\mathbf{M}_{\hat{f}_{1j^*}}$ with respect to $\bar{t}_1, \bar{t}_2, s_1^2,$ and s_2^2 and choose any set of measurement times with these means and variances. We find that $\mathbf{M}_{\hat{f}_{1j^*}}$ can be minimized by collecting data at the following times (see Appendix D):

$$\begin{aligned} \bar{t}_{1\min} &= \frac{t_{1j^*} (\delta_{\alpha_1}^2 ((1-\rho^2)\delta_{\alpha_2}^2 + \sigma_2^2/m_2) + (\sigma_1^2/m_1)(\delta_{\alpha_2}^2 + \sigma_2^2/m_2))}{\delta_{\alpha_1}^2 ((1-\rho^2)\delta_{\alpha_2}^2 + \sigma_2^2/m_2)}, \quad (81) \end{aligned}$$

$$\bar{t}_{2\min} = 0. \quad (82)$$

The optimal mean measurement time for the primary response variable, $\bar{t}_{1\min}$ in (81), lies slightly above the prediction time. In the limit as $m_2 \downarrow 0$, the solution becomes the univariate solution:

$$\bar{t}_{1\min} = \frac{t_{1j^*} (\delta_{\alpha_1}^2 + \sigma_1^2/m_1)}{\delta_{\alpha_1}^2}. \quad (83)$$

Again, the absolute minimum of $\mathbf{M}_{\hat{f}_{1j^*}}$ is not always located in the feasible region defined by $0 \leq t_{rj} \leq T$. More specifically, the absolute minimum of the unconstrained function is located outside of the feasible region if and only if (see Appendix D):

$$\bar{t}_{1\min} \geq T. \quad (84)$$

Under this condition, it seems logical that the minimum would occur if we were to collect all the primary data at time T and all the secondary data at time zero. However, a simulation study similar to that described above revealed counterexamples, which occurred when $\omega > 0.99$. When T was decreased (or the ranges of the variance components for the slope were increased), counterexamples also occurred at smaller values of ω . Therefore, the optimal measurement scheme for the bivariate case appears to depend on ω and T . When ω is small or T is large (in comparison to the magnitudes of δ_{β_1} and δ_{β_2}), the optimal design seems to be collecting all the primary data at $t = T$ and all the secondary data at $t = 0$. When ω is large or T is small (compared to the magnitudes of δ_{β_1} and δ_{β_2}), better designs are likely to be found numerically.

In summary, the average prediction accuracy for the simple linear time-dependent model depends not only on the number of measurements collected, but also on the times when these measurements are taken. In the univariate case, the prediction accuracy depends on these times only through their mean and variance. In the example of the two-process model, the optimal mean of the measurement times is slightly after the prediction time, where the delay increases with more prior information and with fewer or less informative data. When little prior information on the intercept is available, it is usually possible to collect data so

that the absolute minimum of $\mathbf{M}_{\hat{f}_{j^*}}$ is achieved. In the case where the theoretical minimum cannot be achieved (i.e., (66) is not satisfied), minimization of $\mathbf{M}_{\hat{f}_{j^*}}$ can be achieved by collecting all the data at time T .

In the bivariate case of our example, the prediction accuracy depends on the measurement times only through the means and variances of the primary and secondary measurement times. Furthermore, $\mathbf{M}_{\hat{f}_{1j^*}}$ is minimized by centering the primary measurement times above t_{1j^*} (see (81)) as in the univariate case and collecting all secondary measurements at time zero. As in the univariate case, when little prior information on the primary intercept is available, it is usually possible to collect data so that the absolute minimum of $\mathbf{M}_{\hat{f}_{1j^*}}$ is achieved. In the case where this minimum cannot be achieved (i.e., (84) is not satisfied), minimization of $\mathbf{M}_{\hat{f}_{1j^*}}$ can usually be obtained for parameter ranges considered in our simulation by collecting all primary data at time T , and all secondary data at time zero.

9. Nonlinear Models with Time Dependency

Lastly, we focus briefly on the nonlinear case, where the BMSE generally lacks a closed form solution. Obtaining the BMSE of the MMSE estimator for nonlinear models typically requires numerical integration of the joint probability density of \mathbf{y} and $\boldsymbol{\theta}$. We illustrate this with an example in which we numerically estimate the prediction BMSE given a nonlinear model and a single primary response measurement and show how it can be improved by a secondary response measurement.

We consider a two-parameter sinusoidal model of circadian (i.e., 24 h) rhythmicity, defined for a given subject and a bivariate response ($r = 1, 2$), as follows:

$$y_{rj} = f_{rj} + \epsilon_{rj} = A_r \sin\left(\frac{2\pi(t_{rj} - \phi)}{24}\right) + \epsilon_{rj}, \quad (85)$$

where t is in hours; A_r represents a response-specific amplitude; and ϕ represents the phase, which is assumed to be common to both response types. For our example, we assume

$$\begin{aligned} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 5 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 & 0.95 \\ 0.95 & 1 \end{pmatrix}\right), \\ \phi &\sim \mathcal{N}(0, 2). \end{aligned} \quad (86)$$

We simulated $n = 1000$ individuals from this model using normally distributed errors, with primary and secondary response standard deviations of $\sigma_1 = 0.25$ and $\sigma_2 = 0.1$. For each individual, we simulated a single primary response at time $t_{11} = 14$ and a single secondary response at time $t_{12} = 22$. Bayesian forecasting was performed using Markov Chain Monte Carlo (MCMC) with a chain length of 100,000 to obtain the MMSE predictions f_{1j^*} for each individual at time $t_{1j^*} = 24$. Predictions were constructed using a primary data point alone, and also using both a primary data point and a secondary data point.

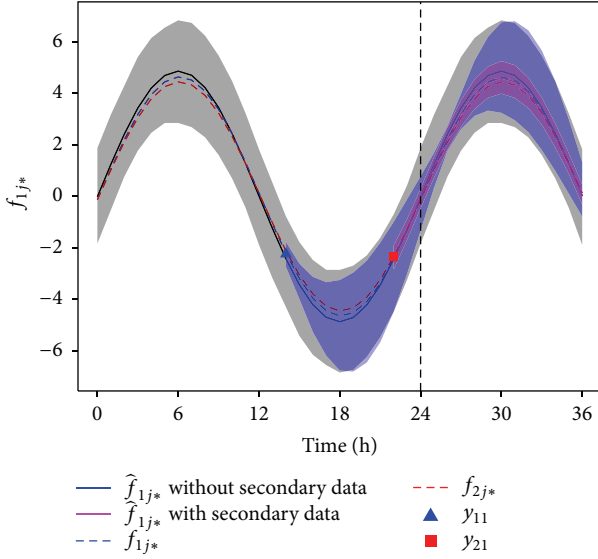


FIGURE 10: MMSE predictions of the expected primary response \hat{f}_{1j^*} and corresponding 95% confidence intervals for a randomly simulated individual from the model specified in (85). The MMSE estimator \hat{f}_{1j^*} is determined assuming each of the following: no data was observed (black line and gray confidence interval), a single primary response was observed (blue line and confidence interval), and both a primary and a secondary response were observed (purple line and confidence interval). The expected primary and secondary responses using the subject's simulated parameter values are shown with blue and red dashed lines. The black vertical dashed line shows the time at which predictions are made.

Figure 10 shows these predictions for a randomly chosen individual. The 95% confidence intervals on f_{1j^*} were constructed using the quantiles of the posterior distribution for f_{1j^*} . For the individual considered, the predictions based on both response variables (purple line and shading) are at many times substantially more accurate (i.e., they have smaller posterior variance) than the predictions based on only the primary response (blue line and shading).

The average accuracy over individuals was assessed at time $t_{1j^*} = 24$ by estimating the BMSE, as follows (cf. (3)):

$$\widehat{\mathbf{M}}_{\hat{f}_{1ij^*}} = \frac{1}{N} \sum_{i=1}^N (f_{1ij^*} - \hat{f}_{1ij^*})^2. \quad (87)$$

The estimated BMSE of \hat{f}_{1ij^*} when using only the primary data points was 0.53, and the estimated BMSE of \hat{f}_{1ij^*} when using both the primary and secondary data points was 0.20. These results suggest that there are nonlinear modeling scenarios where secondary response variables can improve predictions of primary response variables considerably via between-subjects correlation.

10. Discussion

In this paper, we illustrated in a Bayesian, repeated-measures framework how to improve the prediction accuracy of the

expected response, as measured by the BMSE, for a primary response variable in the presence of a secondary response variable that is correlated with the primary response variable between subjects. To set up the general procedure of improving prediction accuracy through Bayesian forecasting, we constructed the BMSE for a simple univariate random intercept model.

We applied the general procedure to a bivariate random intercept model and derived the BMSE of the primary response predictions for this model. We studied how the BMSE depends on the number of observations from the secondary response and found that, for a fixed number of primary response observations, the BMSE is bounded below as specified by (40). The potential value of considering a secondary response may be assessed by considering whether this lower bound represents a meaningful improvement in the expected prediction accuracy.

Assuming the availability of a reasonably highly correlated secondary variable we also addressed the problem of determining the number of primary and secondary measurements needed to obtain a given average accuracy at minimal cost. We derived equations for the solution to this problem and illustrated their use with an example from sleep research. Given previously observed means, variances, and between-subjects correlation for polysomnographic (primary) and actigraphic (secondary) measurements of wakefulness after sleep onset and assumed reasonable measurement costs, we found that, to obtain an average prediction accuracy of 15 min, the use of actigraphy in addition to polysomnography resulted in a substantial reduction in estimated data collection costs as compared to polysomnography alone.

We then considered a steady-state, linear approximation of the homeostatic component for the two-process model of sleep regulation [10, 31] in the univariate case. We found that the minimization of the BMSE with respect to the vector of times at which the data are collected can be divided into two subcases. In one subcase, the BMSE is minimized by collecting data at times with a mean value slightly above the time at which predictions are to be made, with the offset being inversely proportional to the variance of the prior on the intercept. In the second subcase, all of the data is to be collected at the maximal time point. We extended the results of this example to the bivariate case, which again can be divided into two subcases. In one subcase, the BMSE is minimized by collecting the data so that the secondary measurement times have a mean of zero, and the primary variable is collected at times with a mean value somewhat above the time at which the predictions are to be made, like in the univariate case. In the other subcase, the time points that minimize the BMSE may best be found numerically.

Finally, we considered a nonlinear circadian model and determined the improvement in individualized prediction accuracy from a single primary data point versus both a single primary data point and a single secondary data point. For this particular model (85), we found the improvement to be substantial, suggesting that there are cases for nonlinear models where a secondary variable can substantially improve prediction accuracy for the primary variable, given a reasonably high between-subjects correlation.

In conclusion, depending on the between- and within-subjects variance components and the between-subjects correlation between primary and secondary responses, using secondary response data can be effective in increasing the individualized prediction accuracy on the primary response variable in Bayesian forecasting.

This work represents an improvement over the work of Chandler and colleagues [19], who proposed incorporating secondary variables in individualized performance predictions as covariates in a generalized linear model. An advantage of their approach is that it accounts for perturbations on system dynamics from external factors that are common to the outcome variables considered. A drawback is that the approach does not accumulate information about individual differences over time and therefore does not become increasingly accurate for individualized predictions as more data are collected for the individual at hand. Furthermore, the technique requires secondary data to be measured at the same time as the primary data is to be predicted. This is not a requirement for the presently proposed method, for which individualized predictions can be made for any given time, even if secondary data are unavailable then.

That said, here we considered only Bayesian models with diagonally structured error covariance matrices. Such models do not account for correlation within subjects between response types. The work of Chandler and colleagues [19] does account for such correlation, using a fixed linear relationship between primary and secondary responses. We did not consider this possibility here, using merely a diagonal error covariance matrix with one parameter for each response type. However, the Bayesian modeling framework in this paper can be expanded to account for correlation within subjects between response types by adding additional structure to the error covariance matrix.

Finally, the models described here assume that error variance is constant over repeated measures and across different subjects. This constraint can be relaxed easily by allowing the error covariance matrix to have different elements for different individuals.

The multivariate repeated-measures Bayesian forecasting framework presented here may be useful in a variety of clinical settings. One example is modeling the disabling effects of chronic back pain, where pain-related fear may be a good choice for a secondary variable [33]. For other examples, a rich literature in this area can be found in the domain of anesthesiology [34], where clinical applications of multivariate Bayesian forecasting abound.

Appendices

A. Expression for the Univariate Linear Time-Dependent Model

Theorem A.1. *The BMSE of MMSE predictions for prediction BMSE in the univariate, linear, time-dependent model given by (58) is given by*

$$\mathbf{M}_{\hat{f}_{j^*}} = \mathbf{h}' (\mathbf{C}_\theta^{-1} + \mathbf{H}' \mathbf{C}_\epsilon^{-1} \mathbf{H})^{-1} \mathbf{h}, \quad (\text{A.1})$$

where

$$\mathbf{h} = \begin{pmatrix} 1 \\ t_{j^*} \end{pmatrix}, \quad (\text{A.2})$$

where t_{j^*} represents the time for which the expected response is predicted, \mathbf{C}_θ is the prior variance matrix for the parameter vector $\boldsymbol{\theta}$, \mathbf{H} is the design matrix, \mathbf{C}_ϵ is the error covariance matrix, and $\mathbf{M}_{\hat{f}_{j^*}}$ is defined in (3).

Proof. Since

$$f_{j^*} = \mathbf{h}' \boldsymbol{\theta} \quad (\text{A.3})$$

and since the MMSE estimator commutes over linear transformations [23], the MMSE estimator of the expected response is given by

$$\hat{f}_{j^*} = \mathbf{h}' \hat{\boldsymbol{\theta}}. \quad (\text{A.4})$$

Therefore,

$$\begin{aligned} \mathbf{M}_{\hat{f}_{j^*}} &\equiv E \left[(f_{j^*} - \hat{f}_{j^*})^2 \right] \\ &= E \left[(f_{j^*} - \hat{f}_{j^*}) (f_{j^*} - \hat{f}_{j^*})' \right] \\ &= E \left[\mathbf{h}' (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{h} \right] \\ &= \mathbf{h}' E \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \right] \mathbf{h} = \mathbf{h}' \mathbf{M}_{\hat{\boldsymbol{\theta}}} \mathbf{h}, \end{aligned} \quad (\text{A.5})$$

where the parameter BMSE matrix $\mathbf{M}_{\hat{\boldsymbol{\theta}}}$ is given by [23] as follows:

$$\mathbf{M}_{\hat{\boldsymbol{\theta}}} \equiv E \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \right] = (\mathbf{C}_\theta^{-1} + \mathbf{H}' \mathbf{C}_\epsilon^{-1} \mathbf{H})^{-1}. \quad (\text{A.6})$$

Therefore, the BMSE of the estimated response is given by (A.1). \square

Theorem A.2. *The BMSE of MMSE predictions for the univariate, linear, time-dependent model given by (58) is given by*

$$\begin{aligned} \mathbf{M}_{\hat{f}_{j^*}} &= \frac{1/\delta_\beta^2 + ms^2/\sigma^2 + m(\bar{t} - t_{j^*})^2/\sigma^2 + t_{j^*}^2/\delta_\alpha^2}{(1/\delta_\alpha^2 + m/\sigma^2)(1/\delta_\beta^2 + ms^2/\sigma^2) + m\bar{t}^2/\delta_\alpha^2 \sigma^2}, \end{aligned} \quad (\text{A.7})$$

where $\mathbf{M}_{\hat{f}_{j^*}}$ is given in (A.1).

Proof. For this model, the inverse of the between-subjects covariance matrix is given by

$$\mathbf{C}_\theta^{-1} = \begin{pmatrix} \frac{1}{\delta_\alpha^2} & 0 \\ 0 & \frac{1}{\delta_\beta^2} \end{pmatrix}. \quad (\text{A.8})$$

Furthermore, the parameter BMSE matrix is given by (see Theorem A.1)

$$\mathbf{M}_{\hat{f}_{j^*}} = \mathbf{h}' (\mathbf{C}_\theta^{-1} + \mathbf{H}' \mathbf{C}_\epsilon^{-1} \mathbf{H})^{-1} \mathbf{h}, \quad (\text{A.9})$$

where \mathbf{h} is the covariate vector at the time at which we want to make predictions:

$$\mathbf{h} = \begin{pmatrix} 1 \\ t_{j^*} \end{pmatrix}. \quad (\text{A.10})$$

The error variance matrix is as follows:

$$\mathbf{C}_\epsilon = \sigma^2 \mathbf{I}. \quad (\text{A.11})$$

The design matrix is as follows:

$$\mathbf{H} = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix}. \quad (\text{A.12})$$

The matrix multiplication $\mathbf{H}'\mathbf{C}_\epsilon^{-1}\mathbf{H}$ can be computed to be

$$\mathbf{H}'\mathbf{C}_\epsilon^{-1}\mathbf{H} = \frac{1}{\sigma^2} \begin{pmatrix} m & \sum_{j=1}^m t_j \\ \sum_{j=1}^m t_j & \sum_{j=1}^m t_j^2 \end{pmatrix}. \quad (\text{A.13})$$

The matrix inverse $(\mathbf{C}_\theta^{-1} + \mathbf{H}'\mathbf{C}_\epsilon^{-1}\mathbf{H})^{-1}$ can then be computed as follows:

$$\begin{aligned} & (\mathbf{C}_\theta^{-1} + \mathbf{H}'\mathbf{C}_\epsilon^{-1}\mathbf{H})^{-1} \\ &= \begin{pmatrix} \frac{1}{\delta_\alpha^2} + \frac{m}{\sigma^2} & \sum_{j=1}^m \frac{t_j}{\sigma^2} \\ \sum_{j=1}^m \frac{t_j}{\sigma^2} & \frac{1}{\delta_\beta^2} + \frac{\sum_{j=1}^m t_j^2}{\sigma^2} \end{pmatrix}^{-1} \\ &= \left(\left(\frac{1}{\delta_\alpha^2} + \frac{m}{\sigma^2} \right) \left(\frac{1}{\delta_\beta^2} + \frac{\sum_{j=1}^m t_j^2}{\sigma^2} \right) - \left(\frac{\sum_{j=1}^m t_j}{\sigma^2} \right)^2 \right)^{-1} \\ &\quad \cdot \begin{pmatrix} \frac{1}{\delta_\beta^2} + \frac{\sum_{j=1}^m t_j^2}{\sigma^2} & -\frac{\sum_{j=1}^m t_j}{\sigma^2} \\ -\frac{\sum_{j=1}^m t_j}{\sigma^2} & \frac{1}{\delta_\alpha^2} + \frac{m}{\sigma^2} \end{pmatrix}. \end{aligned} \quad (\text{A.14})$$

Computing the quadratic form, we find that

$$\mathbf{M}_{\hat{f}_{j^*}} = \frac{1/\delta_\beta^2 + (\sum_{j=1}^m t_j^2)/\sigma^2 - (2t_{j^*} \sum_{j=1}^m t_j)/\sigma^2 + t_{j^*}^2 (1/\delta_\alpha^2 + m/\sigma^2)}{(1/\delta_\alpha^2 + m/\sigma^2) (1/\delta_\beta^2 + (\sum_{j=1}^m t_j^2)/\sigma^2) - ((\sum_{j=1}^m t_j)/\sigma^2)^2}. \quad (\text{A.15})$$

We note the following decomposition of the sum of squared times:

$$\begin{aligned} \sum_{j=1}^m t_j^2 &= \sum_{j=1}^m (t_j - \bar{t} + \bar{t})^2 = \sum_{j=1}^m (t_j - \bar{t})^2 + \sum_{j=1}^m (\bar{t})^2 \\ &= m(s^2 + \bar{t}^2). \end{aligned} \quad (\text{A.16})$$

Including this decomposition, we find that

$$\begin{aligned} \mathbf{M}_{\hat{f}_{j^*}} &= \frac{1/\delta_\beta^2 + ms^2/\sigma^2 + m\bar{t}^2/\sigma^2 - 2mt_{j^*}\bar{t}/\sigma^2 + t_{j^*}^2/\delta_\alpha^2 + mt_{j^*}^2/\sigma^2}{(1/\delta_\alpha^2 + m/\sigma^2) (1/\delta_\beta^2 + m(s^2 + \bar{t}^2)/\sigma^2) - (m\bar{t}/\sigma^2)^2} \\ &= \frac{1/\delta_\beta^2 + ms^2/\sigma^2 + (m/\sigma^2) (\bar{t}^2 - 2t_{j^*}\bar{t} + t_{j^*}^2) + t_{j^*}^2/\delta_\alpha^2}{(1/\delta_\alpha^2 + m/\sigma^2) (1/\delta_\beta^2 + m(s^2 + \bar{t}^2)/\sigma^2) - (m\bar{t}/\sigma^2)^2} \\ &= \frac{1/\delta_\beta^2 + ms^2/\sigma^2 + m(\bar{t} - t_{j^*})^2/\sigma^2 + t_{j^*}^2/\delta_\alpha^2}{(1/\delta_\alpha^2 + m/\sigma^2) (1/\delta_\beta^2 + ms^2/\sigma^2) + m\bar{t}^2/\delta_\alpha^2 \sigma^2}. \end{aligned} \quad (\text{A.17})$$

□

B. Unconstrained Minimization of the BMSE for the Univariate, Time-Dependent Linear Model

Theorem B.1. For the univariate, linear, time-dependent model given by (58), $\mathbf{M}_{\hat{f}_{j^*}}$ as given by (A.17) exhibits an absolute minimum at a point $(\bar{t}_{\min}, s_{\min}^2)$ if and only if

$$\mathbf{M}_{\hat{f}_{j^*}} = \frac{1}{1/\delta_\alpha^2 + m/\sigma^2}. \quad (\text{B.1})$$

Proof. Recall that \bar{t} and s^2 are defined by (63) and (64). To prevent ambiguous notation, let $\gamma = s^2$. Consider the derivative of $\mathbf{M}_{\hat{f}_{j^*}}$ with respect to γ :

$$\begin{aligned} & \frac{\partial \mathbf{M}_{\hat{f}_{j^*}}(\bar{t}, \gamma)}{\partial \gamma} \\ &= - \frac{m\sigma^2 (t_{j^*}\sigma^2 + m(-\bar{t} + t_{j^*})\delta_\alpha^2)^2 \delta_\beta^4}{(m\delta_\alpha^2 (\sigma^2 + m\gamma\delta_\beta^2) + \sigma^2 (\sigma^2 + m(\gamma + \bar{t}^2)\delta_\beta^2))^2}. \end{aligned} \quad (\text{B.2})$$

We can represent this derivative more simply as follows:

$$\frac{\partial \mathbf{M}_{\hat{f}_j^*}(\bar{t}, \gamma)}{\partial \gamma} = -\frac{m\sigma^2 c_1^2 \delta_\beta^4}{c_2^2}, \quad (\text{B.3})$$

where the values of c_1 and c_2 are found explicitly from (B.2). It is easy to show that $c_1^2 \geq 0$ and $c_2^2 > 0$, and consequently

$$\frac{\partial \mathbf{M}_{\hat{f}_j^*}(\bar{t}, \gamma)}{\partial \gamma} \leq 0. \quad (\text{B.4})$$

Therefore, for a fixed \bar{t} , $\mathbf{M}_{\hat{f}_j^*}$ either decreases or remains constant as we increase γ . It follows that, for fixed value of \bar{t} , $\mathbf{M}_{\hat{f}_j^*}$ is bounded below by $\mathbf{M}_{\hat{f}_j^*}$ in the limit as $\gamma \rightarrow \infty$:

$$\lim_{\gamma \rightarrow \infty} \mathbf{M}_{\hat{f}_j^*}(\bar{t}, \gamma) = \frac{1}{1/\delta_\alpha^2 + m/\sigma^2}. \quad (\text{B.5})$$

$$\frac{\partial \mathbf{M}_{\hat{f}_j^*}(\bar{t}, s^2)}{\partial \bar{t}} = \frac{2m\sigma^2 (t_{j^*}\sigma^2 + m(-\bar{t} + t_{j^*})\delta_\alpha^2)\delta_\beta^2 (\bar{t}t_{j^*}\sigma^2\delta_\beta^2 + \delta_\alpha^2(\sigma^2 + ms^2\delta_\beta^2))}{(m\delta_\alpha^2(\sigma^2 + ms^2\delta_\beta^2) + \sigma^2(\sigma^2 + m(s^2 + \bar{t}^2)\delta_\beta^2))^2}. \quad (\text{B.8})$$

The derivative with respect to s^2 is given by (B.2). Setting both derivatives equal to zero, we find the following critical point:

$$\bar{t}_{\min} = \frac{t_{j^*}(\sigma^2/m + \delta_\alpha^2)}{\delta_\alpha^2}. \quad (\text{B.9})$$

Evaluating the BMSE at this critical point, we find that

$$\mathbf{M}_{\hat{f}_j^*}(\bar{t}_{\min}, s^2) = \frac{1}{1/\delta_\alpha^2 + m/\sigma^2}, \quad (\text{B.10})$$

which is sufficient to show that $\mathbf{M}_{\hat{f}_j^*}$ exhibits an absolute minimum at this point.

Given the constraint that $0 \leq t_j \leq T$ for each j , \bar{t}_{\min} is a feasible point only when $0 \leq \bar{t}_{\min} \leq T$. It is easy to see that $\bar{t}_{\min} > 0$ in all cases. Therefore, the absolute minimum can

$$\mathbf{M}_{\hat{f}_j^*} = \frac{(t_1^2 + t_2^2)/\sigma^2 - (2(t_1 + t_2)t_{j^*})/\sigma^2 + t_{j^*}^2((2/\sigma^2) + (1/\delta_\alpha^2)) + (1/\delta_\beta^2)}{-(t_1 + t_2)^2/\sigma^4 + ((2/\sigma^2) + (1/\delta_\alpha^2))((t_1^2 + t_2^2)/\sigma^2 + (1/\delta_\beta^2))}. \quad (\text{C.2})$$

From Theorem B.2 we know that there are no critical points inside the feasible region, and the solution must therefore exist on the boundary of the region, which is defined by the following line segments:

Since the limit of $\mathbf{M}_{\hat{f}_j^*}$ does not depend on the values of \bar{t} and γ , any values of \bar{t} and γ for which $\mathbf{M}_{\hat{f}_j^*}$ takes on the value given by (B.5) are an absolute minimum of $\mathbf{M}_{\hat{f}_j^*}$. \square

Theorem B.2. For the univariate, linear, time-dependent model given by (58), an absolute minimum of $\mathbf{M}_{\hat{f}_j^*}$ as given by (A.17), inside the region defined by $0 \leq t_j \leq T$, can be obtained by collecting data at times t_j such that

$$\bar{t} = \frac{t_{j^*}(\sigma^2/m + \delta_\alpha^2)}{\delta_\alpha^2}, \quad (\text{B.6})$$

if and only if

$$\bar{t} \leq T. \quad (\text{B.7})$$

Proof. The critical points of the unconstrained function are found by taking the derivative of $\mathbf{M}_{\hat{f}_j^*}$ with respect to \bar{t} and s^2 and setting these derivatives to zero. The derivative with respect to \bar{t} is as follows:

be achieved by collecting all data at \bar{t}_{\min} if and only if $\bar{t}_{\min} \leq T$. \square

C. Constrained Minimization of the BMSE for the Univariate, Time-Dependent Linear Model with Two Time Points

Theorem C.1. For the univariate, linear, time-dependent model given by (58) with $m = 2$ and $0 \leq t_j \leq T$, if

$$\frac{t_{j^*}(\sigma^2/2 + \delta_\alpha^2)}{\delta_\alpha^2} > T, \quad (\text{C.1})$$

then $\mathbf{M}_{\hat{f}_j^*}$ as given by (A.17) is minimized by collecting data at times such that $t_1 = t_2 = T$.

Proof. Substituting $m = 2$ into (A.15) we find that

$$A: t_1 \in [0, T], \quad t_2 = 0;$$

$$B: t_1 = 0, \quad t_2 \in [0, T];$$

$$\begin{aligned} C: t_1 &\in [0, T], \quad t_2 = T; \\ D: t_1 &= T, \quad t_2 \in [0, T]. \end{aligned} \quad (C.3)$$

Note that $\mathbf{M}_{\hat{f}_{j^*}}$ is symmetric in t_j :

$$\mathbf{M}_{\hat{f}_{j^*}}(t_1 = a, t_2 = b) = \mathbf{M}_{\hat{f}_{j^*}}(t_1 = b, t_2 = a). \quad (C.4)$$

Therefore, it is sufficient to consider only line segments A and C. We first consider finding the minimum value on line segment A. Setting $t_2 = 0$ and then taking the derivative of $\mathbf{M}_{\hat{f}_{j^*}}$ with respect to t_1 , and solving for t_1 , we find the following two critical points:

$$t_{1\min}^{A+} = \frac{t_{j^*}(\sigma^2 + 2\delta_\alpha^2)}{\delta_\alpha^2}, \quad (C.5)$$

$$t_{1\min}^{A-} = -\frac{\sigma^2\delta_\alpha^2}{t_{j^*}(\sigma^2 + \delta_\alpha^2)\delta_\beta^2}. \quad (C.6)$$

Applying condition (C.1) to $t_{1\min}^{A+}$, we find that $t_{1\min}^{A+} > 2T$. Taking the second derivative of $\mathbf{M}_{\hat{f}_{j^*}}$ with respect to t_1 , such that $t_2 = 0$, and evaluating it at $t_{1\min}^{A+}$ result in

$$\begin{aligned} \frac{\partial^2 \mathbf{M}_{\hat{f}_{j^*}}}{\partial t_1^2} & \\ &= \frac{2\sigma^2\delta_\alpha^8\delta_\beta^2}{(\sigma^2 + 2\delta_\alpha^2)^2 \left(\sigma^4 t_{j^*}^2 \delta_\beta^2 + 3\sigma^2 t_{j^*} \delta_\alpha^2 \delta_\beta^2 + \delta_\alpha^4 (\sigma^2 + 2t_{j^*}^2 \delta_\beta^2) \right)}, \end{aligned} \quad (C.7)$$

which is positive. Therefore (C.5) represents a minimum.

$$\frac{\partial^2 \mathbf{M}_{\hat{f}_{j^*}}}{\partial t_1^2} = \frac{2\sigma^2\delta_\alpha^8\delta_\beta^2}{(\sigma^2 + 2\delta_\alpha^2)^2 \left(\sigma^4 t_{j^*}^2 \delta_\beta^2 + \sigma^2 t_{j^*} (-2t_f + 3t_{j^*}) \delta_\alpha^2 \delta_\beta^2 + \delta_\alpha^4 (\sigma^2 + 2t_f^2 \delta_\beta^2 - 4t_f t_{j^*} \delta_\beta^2 + 2t_{j^*}^2 \delta_\beta^2) \right)}. \quad (C.10)$$

Evaluating (C.10) at $t_{j^*} = T$ results in

$$\frac{\partial^2 \mathbf{M}_{\hat{f}_{j^*}}}{\partial t_1^2} = \frac{2\delta_\alpha^8\delta_\beta^2}{(\sigma^2 + 2\delta_\alpha^2)^2 \left(\delta_\alpha^4 + \sigma^2 t_f^2 \delta_\beta^2 + t_f^2 \delta_\alpha^2 \delta_\beta^2 \right)}, \quad (C.11)$$

It is easy to see that $t_{1\min}^{A-}$ is always negative. Taking the second derivative of $\mathbf{M}_{\hat{f}_{j^*}}$ with respect to t_1 , such that $t_2 = 0$, and evaluating it at $t_{1\min}^{A-}$ (C.6) result in

$$\begin{aligned} \frac{\partial^2 \mathbf{M}_{\hat{f}_{j^*}}}{\partial t_1^2} & \\ &= -\frac{2t_n^4 (\sigma^2 + \delta_\alpha^2)^2 \delta_\beta^6}{\sigma^6 t_n^2 \delta_\beta^2 + 3\sigma^4 t_n^2 \delta_\alpha^2 \delta_\beta^2 + \delta_\alpha^4 (\sigma^4 + 2\sigma^2 t_n^2 \delta_\beta^2)}, \end{aligned} \quad (C.8)$$

which is negative. Therefore $t_{1\min}^{A-}$ represents a maximum.

Since the maximum occurs at a value $t_{1\min}^{A-} < 0$, the minimum occurs at a value $t_{1\min}^{A+} > T$, and these are the only two critical points; the BMSE must be decreasing over the line segment, and the minimum BMSE on the line segment therefore occurs at $t_1 = T, t_2 = 0$.

We next consider finding the minimum value on the line segment C. Setting $t_2 = T$ and then taking the derivative of $\mathbf{M}_{\hat{f}_{j^*}}$ with respect to t_1 , and solving for t_1 , we find the following two critical points:

$$t_{1\min}^{C+} = \frac{\sigma^2 t_{j^*} - T\delta_\alpha^2 + 2t_{j^*}\delta_\alpha^2}{\delta_\alpha^2}, \quad (C.9)$$

$$t_{1\min}^{C-} = \frac{\delta_\alpha^2 (\sigma^2 + T^2 \delta_\beta^2 - T t_{j^*} \delta_\beta^2)}{(-\sigma^2 t_{j^*} + T\delta_\alpha^2 - t_{j^*} \delta_\alpha^2) \delta_\beta^2}.$$

We see that $t_{1\min}^{C+}$ is minimized by applying the lower bound for t_{j^*} given by (C.1), and we find that $t_{1\min}^{C+} > T$. Concerning $t_{1\min}^{C-}$, the numerator can be minimized by applying the upper bound for t_{j^*} of T . In this case, the numerator is $\delta_\alpha^2 \sigma^2$, which is always positive. The maximum value of the denominator is found by substituting in the minimum value for t_{j^*} given by (C.1). We find that the maximum value is $-T\sigma^2 \delta_\alpha^2 \delta_\beta^2 / (\sigma^2 + 2\delta_\alpha^2)$, which is negative. Therefore $t_{1\min}^{C-}$ is always negative.

Taking the second derivative of $\mathbf{M}_{\hat{f}_{j^*}}$ with respect to t_1 , such that $t_2 = T$, and evaluating it at $t_{1\min}^{C+}$ result in

which is positive. In general, (C.10) is positive when

$$\begin{aligned} \sigma^4 t_{j^*}^2 \delta_\beta^2 + \sigma^2 t_{j^*} (-2t_f + 3t_{j^*}) \delta_\alpha^2 \delta_\beta^2 & \\ + \delta_\alpha^4 (\sigma^2 + 2t_f^2 \delta_\beta^2 - 4t_f t_{j^*} \delta_\beta^2 + 2t_{j^*}^2 \delta_\beta^2) & > 0. \end{aligned} \quad (C.12)$$

Setting the expression on the left-hand side to zero and solving for t_{j^*} , we find the following roots:

$$t_{j^*} = \frac{\sigma^2 t_f \delta_\alpha^2 \delta_\beta^2 + 2t_f \delta_\alpha^4 \delta_\beta^2 \pm \sqrt{-\sigma^6 \delta_\alpha^4 \delta_\beta^2 - 3\sigma^4 \delta_\alpha^6 \delta_\beta^2 - 2\sigma^2 \delta_\alpha^8 \delta_\beta^2 - \sigma^4 t_f^2 \delta_\alpha^4 \delta_\beta^4 - 2\sigma^2 t_f^2 \delta_\alpha^6 \delta_\beta^4}}{\sigma^4 \delta_\beta^2 + 3\sigma^2 \delta_\alpha^2 \delta_\beta^2 + 2\delta_\alpha^4 \delta_\beta^2}. \quad (\text{C.13})$$

These roots are imaginary, which implies that (C.10) is always positive, and therefore, $t_{1\min}^{C+}$ is a minimum. Given that the BMSE must always be positive, and there are only two critical points, this implies that $t_{1\min}^{C-}$ is a maximum. Since the maximum occurs at a value of $t_{1\min}^{C-} < 0$ and the minimum occurs at a value of $t_{1\min}^{C+} > T$, the minimum on the line segment must occur at the value $t_1 = T$.

In summary, the minimum value on the line segment A occurs at $t_1 = T$, $t_2 = 0$. Applying symmetry, the BMSE at this point is the same as the BMSE at the points $t_1 = 0$, $t_2 = T$, which is the maximum point for the line segment C. Therefore, the minimum BMSE does not occur on the line segment A. Again, by symmetry, this means that the minimum BMSE does not occur on the line segment B. For the line segment C, the minimum BMSE occurs at $t_1 = T$, $t_2 = T$. Applying symmetry, the minimum BMSE on the line segment D occurs at $t_1 = T$, $t_2 = T$. Since these two points are the same, we find that the overall minimum BMSE within the feasible region occurs at the point $t_1 = t_2 = T$. \square

D. Unconstrained Minimization of the BMSE for the Bivariate, Time-Dependent Linear Model

Theorem D.1. For the bivariate, linear, time-dependent model given by (70), $\mathbf{M}_{\hat{f}_{1j^*}}$ as given by (74) exhibits an absolute minimum at a point $(\bar{t}_{1\min}, \bar{t}_{2\min}, s_{1\min}^2, s_{2\min}^2)$ if and only if

$$\mathbf{M}_{\hat{f}_{1j^*}} = \frac{\delta_{\alpha_1}^2 (\sigma_1^2/m_1) ((1-\rho^2) \delta_{\alpha_2}^2 + \sigma_2^2/m_2)}{(\sigma_1^2/m_1) (\delta_{\alpha_2}^2 + \sigma_2^2/m_2) + a_1^2 ((1-\rho^2) \delta_{\alpha_2}^2 + \sigma_2^2/m_2)}. \quad (\text{D.1})$$

Proof. Recall that \bar{t}_1 , \bar{t}_2 , s_1^2 , and s_2^2 are defined in (77), (78), (79), and (80). To prevent ambiguous notation, let $\gamma_1 = s_1^2$ and $\gamma_2 = s_2^2$. Consider the derivative of $\mathbf{M}_{\hat{f}_{1j^*}}$ (see (74)) with respect to γ_1 :

$$\frac{\partial \mathbf{M}_{\hat{f}_{1j^*}}(\bar{t}_1, \bar{t}_2, \gamma_1, \gamma_2)}{\partial \gamma_1} = -\frac{\delta_{\beta_1}^2 m_1 \sigma_1^2 c_1^2}{c_2^2}, \quad (\text{D.2})$$

where

$$c_1 = \bar{t}_2 \rho \omega \delta_{\alpha_1} \delta_{\alpha_2} \delta_{\beta_2} m_2 \sigma_1^2 \sigma_2^2 + t_{j^*} \delta_{\beta_1} \sigma_1^2 (\delta_{\alpha_2}^2 m_2 (s_2^2 (-1 + \omega^2)) \delta_{\beta_2}^2 m_2 - \sigma_2^2)$$

$$- \sigma_2^2 \left(- (s_2^2 + \bar{t}_2^2) (-1 + \omega^2) \delta_{\beta_2}^2 m_2 + \sigma_2^2 \right) + (\bar{t}_1 - t_{j^*}) \delta_{\alpha_1}^2 \delta_{\beta_1} m_1 \left((-1 + \rho^2) \delta_{\alpha_2}^2 m_2 (s_2^2 (-1 + \omega^2)) \cdot \delta_{\beta_2}^2 m_2 - \sigma_2^2 \right) + \sigma_2^2 \left(- (s_2^2 + \bar{t}_2^2) (-1 + \omega^2) \delta_{\beta_2}^2 m_2 + \sigma_2^2 \right),$$

$$c_2 = -2\bar{t}_1 \bar{t}_2 \rho \omega \delta_{\alpha_1} \delta_{\alpha_2} \delta_{\beta_1} \delta_{\beta_2} m_1 m_2 \sigma_1^2 \sigma_2^2 + \sigma_1^2 \left(\delta_{\alpha_2}^2 m_2 \left(- (s_1^2 + \bar{t}_1^2) \cdot \delta_{\beta_1}^2 m_1 (s_2^2 (-1 + \omega^2)) \delta_{\beta_2}^2 m_2 - \sigma_2^2 \right) + \sigma_1^2 (s_2^2 \delta_{\beta_2}^2 m_2 + \sigma_2^2) \right) + \sigma_2^2 \left(- (s_1^2 + \bar{t}_1^2) \cdot \delta_{\beta_1}^2 m_1 \left((s_2^2 + \bar{t}_2^2) (-1 + \omega^2) \delta_{\beta_2}^2 m_2 - \sigma_2^2 \right) + \sigma_1^2 \left((s_2^2 + \bar{t}_2^2) \delta_{\beta_2}^2 m_2 + \sigma_2^2 \right) \right) + a_1^2 m_1 \left((-1 + \rho^2) \delta_{\alpha_2}^2 m_2 (s_1^2 \delta_{\beta_1}^2 m_1 (s_2^2 (-1 + \omega^2)) \delta_{\beta_2}^2 m_2 - \sigma_2^2) - \sigma_1^2 (s_2^2 \delta_{\beta_2}^2 m_2 + \sigma_2^2) \right) + \sigma_2^2 \left(\sigma_1^2 \left((s_2^2 + \bar{t}_2^2) \delta_{\beta_2}^2 m_2 + \sigma_2^2 \right) + s_1^2 \delta_{\beta_1}^2 m_1 \left(- (s_2^2 + \bar{t}_2^2) (-1 + \omega^2) \delta_{\beta_2}^2 m_2 + \sigma_2^2 \right) \right) + \sigma_2^2 \left(\sigma_1^2 \left((s_2^2 + \bar{t}_2^2) \delta_{\beta_2}^2 m_2 + \sigma_2^2 \right) \right). \quad (\text{D.3})$$

It is easy to show that $c_1^2 \geq 0$ and $c_2^2 \geq 0$, and consequently

$$\frac{\partial \mathbf{M}_{\hat{f}_{1j^*}}(\bar{t}_1, \bar{t}_2, \gamma_1, \gamma_2)}{\partial \gamma_1} \leq 0. \quad (\text{D.4})$$

Consider also the derivative of $\mathbf{M}_{\hat{f}_{1j^*}}$ with respect to γ_2 :

$$\frac{\partial \mathbf{M}_{\hat{f}_{1j^*}}(\bar{t}_1, \bar{t}_2, \gamma_1, \gamma_2)}{\partial \gamma_2} = -\frac{\delta_{\beta_2}^2 m_2 \sigma_1^4 \sigma_2^2 c_3^2}{c_4^2}, \quad (\text{D.5})$$

where

$$c_3 = \bar{t}_2 \rho \delta_{\alpha_1} \delta_{\alpha_2} \delta_{\beta_2} m_2 \left((s_1^2 + \bar{t}_1 (\bar{t}_1 - t_{j^*})) (-1 + \omega^2) \cdot \delta_{\beta_1}^2 m_1 - \sigma_1^2 \right) + (\bar{t}_1 - t_{j^*}) \omega \delta_{\alpha_1}^2 \delta_{\beta_1} m_1 \left((-1 + \rho^2) \cdot \delta_{\alpha_2}^2 m_2 - \sigma_2^2 \right) + t_{j^*} \omega \delta_{\beta_1} \sigma_1^2 (\delta_{\alpha_2}^2 m_2 + \sigma_2^2),$$

$$\begin{aligned}
c_4 = & -2\bar{t}_1\bar{t}_2\rho\omega\delta_{\alpha_1}\delta_{\alpha_2}\delta_{\beta_1}\delta_{\beta_2}m_1m_2\sigma_1^2\sigma_2^2 \\
& + \sigma_1^2\left(\delta_{\alpha_2}^2m_2\left(-\left(s_1^2+\bar{t}_1^2\right)\right.\right. \\
& \cdot \delta_{\beta_1}^2m_1\left.\left(s_2^2(-1+\omega^2)\delta_{\beta_2}^2m_2-\sigma_2^2\right)\right. \\
& \left.+\sigma_1^2\left(s_2^2\delta_{\beta_2}^2m_2+\sigma_2^2\right)\right)+\sigma_2^2\left(-\left(s_1^2+\bar{t}_1^2\right)\right. \\
& \cdot \delta_{\beta_1}^2m_1\left.\left(\left(s_2^2+\bar{t}_2^2\right)(-1+\omega^2)\delta_{\beta_2}^2m_2-\sigma_2^2\right)\right. \\
& \left.+\sigma_1^2\left(\left(s_2^2+\bar{t}_2^2\right)\delta_{\beta_2}^2m_2+\sigma_2^2\right)\right)+a_1^2m_1\left(\left(-1\right.\right. \\
& \left.+\rho^2\right)\delta_{\alpha_2}^2m_2\left(s_1^2\delta_{\beta_1}^2m_1\left(s_2^2(-1+\omega^2)\delta_{\beta_2}^2m_2-\sigma_2^2\right)\right. \\
& \left.-\sigma_1^2\left(s_2^2\delta_{\beta_2}^2m_2+\sigma_2^2\right)\right)+\sigma_2^2\left(\sigma_1^2\left(\left(s_2^2+\bar{t}_2^2\right)\delta_{\beta_2}^2m_2\right.\right. \\
& \left.+\sigma_2^2\right)+s_1^2\delta_{\beta_1}^2m_1\left(-\left(s_2^2+\bar{t}_2^2\right)(-1+\omega^2)\delta_{\beta_2}^2m_2\right. \\
& \left.+\sigma_2^2\right)\left.\right). \tag{D.6}
\end{aligned}$$

It can be shown that $c_3^2 \geq 0$ and $c_4^2 > 0$, and consequently

$$\frac{\partial \mathbf{M}_{\hat{f}_{1j^*}}(\bar{t}_1, \bar{t}_2, \gamma_1, \gamma_2)}{\partial \gamma_2} \leq 0. \tag{D.7}$$

Therefore, for fixed \bar{t}_1 and \bar{t}_2 , $\mathbf{M}_{\hat{f}_{1j^*}}$ either decreases or remains constant as we increase γ_1 or γ_2 . It follows that, for fixed values of \bar{t}_1 and \bar{t}_2 , $\mathbf{M}_{\hat{f}_{1j^*}}$ is bounded below by $\mathbf{M}_{\hat{f}_{1j^*}}$ in the limit as $\gamma_1, \gamma_2 \rightarrow \infty$:

$$\begin{aligned}
& \lim_{\gamma_1, \gamma_2 \rightarrow \infty} \mathbf{M}_{\hat{f}_{1j^*}}(\bar{t}_1, \bar{t}_2, \gamma_1, \gamma_2) \\
& = \frac{\delta_{\alpha_1}^2(\sigma_1^2/m_1)((1-\rho^2)\delta_{\alpha_2}^2+\sigma_2^2/m_2)}{(\sigma_1^2/m_1)(\delta_{\alpha_2}^2+\sigma_2^2/m_2)+a_1^2((1-\rho^2)\delta_{\alpha_2}^2+\sigma_2^2/m_2)}. \tag{D.8}
\end{aligned}$$

Since the limit of $\mathbf{M}_{\hat{f}_{1j^*}}$ does not depend on \bar{t}_1 or \bar{t}_2 , any values of $\bar{t}_1, \bar{t}_2, \gamma_1$, and γ_2 for which $\mathbf{M}_{\hat{f}_{1j^*}}$ takes on the value given by (D.8) are an absolute minimum of $\mathbf{M}_{\hat{f}_{1j^*}}$. \square

Theorem D.2. For the bivariate, linear, time-dependent model given by (70), an absolute minimum of $\mathbf{M}_{\hat{f}_{1j^*}}$ as given by (74), inside the region defined by $0 \leq t_{rj} \leq T$, can be obtained by collecting data at times t_{rj} such that

$$\bar{t}_{1\min} = \frac{t_{1j^*}(\delta_{\alpha_1}^2((1-\rho^2)\delta_{\alpha_2}^2+(\sigma_2^2/m_2))+(\sigma_1^2/m_1)(\delta_{\alpha_2}^2+(\sigma_2^2/m_2)))}{\delta_{\alpha_1}^2((1-\rho^2)\delta_{\alpha_2}^2+(\sigma_2^2/m_2))}, \tag{D.9}$$

$$\bar{t}_{2\min} = 0,$$

if and only if

$$\bar{t}_{1\min} \leq T. \tag{D.10}$$

Proof. Substituting (D.9) into $\mathbf{M}_{\hat{f}_{1j^*}}$, we find that

$$\begin{aligned}
& \mathbf{M}_{\hat{f}_{1j^*}}(\bar{t}_{1\min}, \bar{t}_{2\min}, s_1^2, s_2^2) \\
& = \frac{\delta_{\alpha_1}^2(\sigma_1^2/m_1)((1-\rho^2)\delta_{\alpha_2}^2+(\sigma_2^2/m_2))}{(\sigma_1^2/m_1)(\delta_{\alpha_2}^2+\sigma_2^2/m_2)+a_1^2((1-\rho^2)\delta_{\alpha_2}^2+\sigma_2^2/m_2)}, \tag{D.11}
\end{aligned}$$

which is sufficient to show that any point $(\bar{t}_{1\min}, \bar{t}_{2\min}, s_1^2, s_2^2)$ is an absolute minimum of the unconstrained function (see Theorem D.1).

Given the constraint that $0 \leq t_{1j} \leq T$ for each j , $\bar{t}_{1\min}$ is a feasible point only when $0 \leq \bar{t}_{1\min} \leq T$. It is easy to see that $\bar{t}_{1\min} > 0$ in all cases. Therefore, the absolute minimum can be achieved by collecting all data at $\bar{t}_{1\min}$ if and only if $\bar{t}_{1\min} \leq T$. \square

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors are grateful to Hongbo Dong for help with some of the mathematical proofs in the paper. This research was supported by ONR Grant N00014-13-1-0302 and in part by FMCSA Award DTMC75-07-D-00006 and FAA Award DTFAAC-11-A-00003.

References

- [1] P. D. Bliese and R. E. Ployhart, "Growth modeling using random coefficient models: model building, testing, and illustrations," *Organizational Research Methods*, vol. 5, no. 4, pp. 362–387, 2002.
- [2] E. Olofson, D. F. Dinges, and H. P. A. Van Dongen, "Nonlinear mixed-effects modeling: individualization and prediction," *Aviation, Space, and Environmental Medicine*, vol. 75, no. 3, pp. A134–A140, 2004.
- [3] H. P. A. Van Dongen, G. Maislin, and D. F. Dinges, "Dealing with inter-individual differences in the temporal dynamics of fatigue and performance: importance and techniques," *Aviation, Space, and Environmental Medicine*, vol. 75, no. 3, pp. A147–A154, 2004.
- [4] G. Belenky, N. J. Wesensten, D. R. Thorne et al., "Patterns of performance degradation and restoration during sleep restriction

- and subsequent recovery: a sleep dose-response study,” *Journal of Sleep Research*, vol. 12, no. 1, pp. 1–12, 2003.
- [5] H. P. A. Van Dongen, G. Maislin, J. M. Mullington, and D. F. Dinges, “The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation,” *Sleep*, vol. 26, no. 2, pp. 117–126, 2003.
 - [6] H. P. A. Van Dongen, M. D. Baynard, G. Maislin, and D. F. Dinges, “Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability,” *Sleep*, vol. 27, no. 3, pp. 423–433, 2004.
 - [7] H. P. A. Van Dongen, C. G. Mott, J.-K. Huang, D. J. Mollicone, F. D. McKenzie, and D. F. Dinges, “Optimization of biomathematical model predictions for cognitive performance impairment in individuals: accounting for unknown traits and uncertain states in homeostatic and circadian processes,” *Sleep*, vol. 30, no. 9, pp. 1129–1143, 2007.
 - [8] E. Olofsen, H. P. A. Van Dongen, C. G. Mott, T. J. Balkin, and D. Terman, “Current approaches and challenges to development of an individualized sleep and performance prediction model,” *Open Sleep Journal*, vol. 3, no. 1, pp. 24–43, 2010.
 - [9] H. P. A. Van Dongen, C. G. Mott, J.-K. Huang, D. J. Mollicone, F. D. McKenzie, and D. F. Dinges, “Confidence intervals for individualized performance models,” *Sleep*, vol. 30, no. 9, p. 1083, 2007.
 - [10] A. A. Borbély and P. Achermann, “Sleep homeostasis and models of sleep regulation,” *Journal of Biological Rhythms*, vol. 14, no. 6, pp. 557–568, 1999.
 - [11] S. Daan, D. G. Beersma, and A. A. Borbély, “Timing of human sleep: recovery process gated by a circadian pacemaker,” *The American Journal of Physiology*, vol. 246, no. 2, pp. R161–R183, 1984.
 - [12] G. C. Reinsel, “Mean squared error properties of empirical bayes estimators in a multivariate random effects general linear model,” *Journal of the American Statistical Association*, vol. 80, no. 391, pp. 642–650, 1985.
 - [13] C. Brochot, B. Bessoud, D. Balvay, C.-A. Cuénod, N. Siauve, and F. Y. Bois, “Evaluation of antiangiogenic treatment effects on tumors’ microcirculation by Bayesian physiological pharmacokinetic modeling and magnetic resonance imaging,” *Magnetic Resonance Imaging*, vol. 24, no. 8, pp. 1059–1067, 2006.
 - [14] L. A. Dalton and E. R. Dougherty, “Application of the Bayesian MMSE estimator for classification error to gene expression microarray data,” *Bioinformatics*, vol. 27, no. 13, pp. 1822–1831, 2011.
 - [15] E. W. Kamen and J. K. Su, *Introduction to Optimal Estimation*, Springer, London, UK, 1999.
 - [16] B. B. Morgan Jr., P. S. Winne, and J. Dugan, “The range and consistency of individual differences in continuous work,” *Human Factors*, vol. 22, no. 3, pp. 331–340, 1980.
 - [17] S. R. Hursh, D. P. Redmond, M. L. Johnson et al., “Fatigue models for applied research in warfighting,” *Aviation, Space, and Environmental Medicine*, vol. 75, no. 3, pp. A44–A53, 2004.
 - [18] P. McCauley, L. V. Kalachev, D. J. Mollicone, S. Banks, D. F. Dinges, and H. P. A. Van Dongen, “Dynamic circadian modulation in a biomathematical model for the effects of sleep and sleep loss on waking neurobehavioral performance,” *Sleep*, vol. 36, no. 12, pp. 1987–1997, 2013.
 - [19] J. F. Chandler, R. D. Arnold, J. B. Phillips, and A. E. Turnmire, “Predicting individual differences in response to sleep loss: application of current techniques,” *Aviation, Space, and Environmental Medicine*, vol. 84, no. 9, pp. 927–937, 2013.
 - [20] D. J. Frey, P. Badia, and K. P. Wright Jr., “Inter- and intra-individual variability in performance near the circadian nadir during sleep deprivation,” *Journal of Sleep Research*, vol. 13, no. 4, pp. 305–315, 2004.
 - [21] A. M. Tucker, D. F. Dinges, and H. P. A. Van Dongen, “Trait interindividual differences in the sleep physiology of healthy young adults,” *Journal of Sleep Research*, vol. 16, no. 2, pp. 170–180, 2007.
 - [22] S. R. Hursh and H. P. A. Van Dongen, “Fatigue and performance modeling,” in *Principles and Practice of Sleep Medicine*, M. H. Kryger, T. Roth, and W. C. Roth, Eds., pp. 745–752, Elsevier Saunders, St. Louis, Mo, USA, 5th edition, 2010.
 - [23] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory: Volume I*, Prentice Hall, Upper Saddle River, NJ, USA, 1993.
 - [24] A. T. M. Van de Water, A. Holmes, and D. A. Hurley, “Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography—a systematic review,” *Journal of Sleep Research*, vol. 20, no. 1, pp. 183–200, 2011.
 - [25] A. Rechtschaffen and A. Kales, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Participants*, Publication no. 204, Neurological Information Network, National Institutes of Health, Bethesda, Md, USA, 1969.
 - [26] R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin, “Automatic sleep/wake identification from wrist activity,” *Sleep*, vol. 15, no. 5, pp. 461–469, 1992.
 - [27] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Development Core Team, “nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-111,” 2013, <https://cran.r-project.org/web/packages/nlme/nlme.pdf>.
 - [28] N. M. Laird and J. H. Ware, “Random-effects models for longitudinal data,” *Biometrics*, vol. 38, no. 4, pp. 963–974, 1982.
 - [29] J. Pinheiro and D. Bates, *Mixed-Effects Models in S and S-PLUS*, Springer, New York, NY, USA, 2009.
 - [30] K. Chaloner, “Optimal Bayesian experimental design for linear models,” *The Annals of Statistics*, vol. 12, no. 1, pp. 283–300, 1984.
 - [31] A. A. Borbély, “A two process model of sleep regulation,” *Human Neurobiology*, vol. 1, no. 3, pp. 195–204, 1982.
 - [32] I. G. Campbell, L. M. Higgins, N. Darchia, and I. Feinberg, “Homeostatic behavior of Fast Fourier Transform power in very low frequency non-rapid eye movement human electroencephalogram,” *Neuroscience*, vol. 140, no. 4, pp. 1395–1399, 2006.
 - [33] G. Crombez, J. W. S. Vlaeyen, P. H. T. G. Heuts, and R. Lysens, “Pain-related fear is more disabling than pain itself: evidence on the role of pain-related fear in chronic back pain disability,” *Pain*, vol. 80, no. 1-2, pp. 329–339, 1999.
 - [34] E. Olofsen and A. Dahan, “Population pharmacokinetics/pharmacodynamics of anesthetics,” *AAPS Journal*, vol. 7, no. 2, pp. E383–E389, 2005.