Appendix for "Bayesian logistic regression for online recalibration and revision of clinical prediction models with guarantees"

Notation	Description
General terms	
T	Time horizon
d	Number of variables
(x_t, y_t)	Observed variables and outcome at time t
$\hat{f}_t : \mathcal{X} \mapsto \mathbb{R}$	Underlying prediction model at time t
$\hat{A}_t : \mathbb{R} \times \mathcal{X} \mapsto [0, 1]$	Model revision deployed at time t: A function that maps the score from an underlying prediction model at time t and patient variables to a probability. In the special case of model recalibration, $\hat{A}_t$ is a function of only the score.
$\hat{\theta}_t$	Parameters for logistic model revision at time t
$\tau = (\tau_1, \tau_2, \cdots, \tau_s)$	Update times for a given sequence of model revisions
Regret	
$\left(\frac{1}{T}\sum_{t=1}^{T}\left[-\log p\left(y_t, \hat{A}_t(\hat{f}_t(x_t), x_t)\right) + \log p\left(y_t, \hat{A}_1(\hat{f}_1(x_t), x_t)\right)\right]\right)_+$	Type I Regret: The average increase in the negative log likelihood when using the online reviser instead of locking the original model
$\left(\frac{1}{T}\sum_{t=1}^{T}\left[-\log p\left(y_t, \hat{A}_t(\hat{f}_t(x_t), x_t)\right) + \log p\left(y_t, A^*_{\tau,t}(\hat{f}_t(x_t), x_t)\right)\right]\right)_+$	Type II τ-Regret: The average increase in the negative log likelihood when using the online reviser versus the oracle model reviser $\{A^*_{\tau,t} : t = 1, ..., T\}$ with update times τ
BLR and MarBLR parameters	
$N(\theta_{\text{init}}, \Sigma_{\text{init}})$	Gaussian prior in BLR and MarBLR for the logistic revision parameter at time $t = 1$
α	Prior probability in MarBLR that the model revision shifts at time t
δ^2	Factor controlling the variance of the MarBLR prior over shifts in the model revision parameters

Table 1: Terminology and notation

A Practical Implementation of BLR and MarBLR

In this manuscript, we implement MarBLR using a Laplace approximation of the logistic posterior and perform Kalman filtering with collapsing [Gordon and Smith, 1990, West and Harrison, 1997]. Because BLR corresponds to MarBLR with $\alpha = 0$ and $\delta^2 = 0$, we use this same procedure to perform approximate Bayesian inference. The Kalman filtering approach is simple and computationally efficient; We describe the steps below. We note that for the special case of BLR, one can also perform posterior inference by sampling Polya-Gamma latent variables [Polson et al., 2013]. This would allow one to perform full Bayesian inference but is significantly more costly in terms of computation time.

We make predictions and update the posterior using the following recursive procedure. The process is initialized with the Gaussian prior for θ_1 with mean θ_{init} and posterior covariance Σ_{init}. Let $D^{(t)}$ denote the observations up to and including time t.

Prediction step. At time t, let the approximation for $\theta_{t-1}|W_{t-1} = w_{t-1}, D^{(t-1)}$ be the Gaussian distribution with mean $\hat{\theta}_{t-1,w_{t-1}}$ and covariance $\hat{\Sigma}_{t-1,w_{t-1}}$. We also assume $\Pr(W_{t-1} = w_{t-1}|D^{(t-1)})$ is known. We generate predictions at time t using the posterior distribution $\theta_t|D^{(t-1)}$, which is a mixture of the distributions

$$\theta_t|W_t = w_t, W_{t-1} = w_{t-1}, D^{(t-1)} \sim N\left(\hat{\theta}_{t-1,w_{t-1}}, (1+\delta^2 w_t)\hat{\Sigma}_{t-1,w_{t-1}}\right) \tag{1}$$

for $w_t, w_{t=1} \in \{0,1\}$ with weights by $\Pr(W_t = w_t|W_{t-1} = w_{t-1})\Pr(W_{t-1} = w_{t-1}|D^{(t-1)})$. Recall that $\Pr(W_t = 1|W_{t-1} = w_{t-1}) = \alpha$ in the MarBLR prior. We predict that $Y = 1$ for a subject x using the posterior mean of $\Pr(Y = 1|X = x)$.

Update step. Next, we observe a new batch of labeled observations and update the posterior. That is, we must perform inference for $\theta_t|D^{(t)}$, which is a mixture of the distributions $\theta_t|W_t = w_t, W_{t-1} = w_{t-1}, D^{(t)}$ with probability weights $\Pr(W_t = w_t, W_{t-1} = w_{t-1}|D^{(t)})$ for $w_t, w_{t-1} \in \{0,1\}$. Let $\tilde{\ell}_t(\theta, w) = \sum_{i=1}^{n} \log p\left(y_{t,i}|z_{t,i}, \theta\right) + \log p(\theta \mid w, D^{(t-1)})$. We approximate the distribution $\theta_t|W_t = w_t, W_{t-1} = w_{t-1}, D^{(t)}$ using a Gaussian distribution with its mean computed using a Newton update

$$\hat{\theta}_{t,w_t,w_{t-1}} = \hat{\theta}_{t-1,w_{t-1}} - \left[\nabla^2_\theta \tilde{\ell}_t\left(\hat{\theta}_{t-1,w_{t-1}}, w_{t-1}\right)\right]^{-1} \nabla_\theta \tilde{\ell}_t\left(\hat{\theta}_{t-1,w_{t-1}}, w_{t-1}\right) \tag{2}$$

and its covariance as

$$\hat{\Sigma}_{t,w_t,w_{t-1}} = \left[\nabla^2_\theta \tilde{\ell}_t(\hat{\theta}_{t-1,w_{t-1}}, w_{t-1})\right]^{-1}.$$

The probability $\Pr(W_t = w_t, W_{t-1} = w_{t-1} \mid D^{(t)})$, which is proportional to

$$\left[\int_{\theta_t} p\left(y_{t,\cdot} \mid z_{t,\cdot}, \theta_t\right) p\left(\theta_t \mid w_t, w_{t-1}, D^{(t-1)}\right) d\theta_t\right] \Pr(W_t|W_{t-1}) \Pr(W_{t-1}|D^{(t-1)}), \tag{3}$$

is approximated using a Laplace approximation for the integral in (3), i.e.

$$2\pi^{d/2}\left|\left\{\nabla^2_\theta \tilde{\ell}(\hat{\theta}_{t,w_t,w_{t-1}})\right\}^{-1}\right|^{1/2} p\left(y_{t,\cdot} \mid z_{t,\cdot}, \hat{\theta}_{t,w_t,w_{t-1}}\right) p\left(\hat{\theta}_{t,w_t,w_{t-1}} \mid D^{(t-1)}\right). \tag{4}$$

Let $\hat{q}_{w_t,w_{t-1}}$ denote the estimated probability. Finally, we approximate the posterior distribution $\theta_t|W_t = w_t, D^{(t)}$ using a single Gaussian distribution by moment-matching [West and Harrison, 1997, Orguner and Demırekler, 2007] with mean and covariance

$$\hat{\theta}_{t,w_t} = \frac{1}{\hat{q}_{w_t,0} + \hat{q}_{w_t,1}}\left(\hat{q}_{w_t,0}\hat{\theta}_{t,w_t,0} + \hat{q}_{w_t,1}\hat{\theta}_{t,w_t,1}\right) \tag{5}$$

$$\hat{\Sigma}_{t,w_t} = \frac{1}{\hat{q}_{w_t,0} + \hat{q}_{w_t,1}}\left(\hat{q}_{w_t,0}\hat{\Sigma}_{t,w_t,0} + \hat{q}_{w_t,1}\hat{\Sigma}_{t,w_t,1}\right). \tag{6}$$

B Online model revision for batched data

In certain settings, it is more convenient and practical for the data stream to be observed in batches of size $n > 1$. Here we discuss the necessary modifications to our framework for analyzing the performance on an online model reviser for batched data. We denote a batch of observations as $\{(x_{t,i}, y_{t,i}) : i = 1, ..., n\}$ and use the notation $a_{t,\cdot}$ to denote the sequence $(a_{t,1}, ..., a_{t,n})$.

We extend the online model reviser to output a probability distribution over all possible outcomes for a batch of observations, i.e. $\hat{A}_t : \mathbb{R}^n \times \mathcal{X}^n \mapsto \Delta_{2^n}$ where Δ_{2^n} is the probability simplex over all possible outcomes $(y_{t,1}, ..., y_{t,n})$. The loss of the online model reviser over the entire time period is then defined as the average negative log likelihood

$$-\frac{1}{nT}\sum_{t=1}^{T} \log p\left(y_{t,\cdot}, \hat{A}_t\left(\hat{f}_t(x_{t,\cdot}), x_{t,\cdot}\right)\right). \tag{7}$$

This theoretical framework allows predictions from the online model reviser to depend on all unlabeled observations $\{x_{t,i} : i = 1, ..., n\}$. By defining regret with respect to (7), we are able to derive Type I and II regret bounds for the batched setting. This is necessary for analyzing BLR and MarBLR because outcomes are not independent given the observations up to time t in the Bayesian framework. In particular, the outcomes are correlated because of the shared (latent) revision parameter θ_t.

Table 2: Descriptive statistics of variables included in the COPD risk prediction model. Continuous variables are summarized by Mean (SD). Binary/ordinal variables are summarized by number of nonzero entries (%).

Variable	
Diagnosed with COPD	2756 (2.55)
Age at encounter	60.31 (18.60)
Medical history	
Asthma	741 (0.69)
Bronchitis	5855 (5.42)
COPD	14950 (13.84)
Smoking	42651 (39.49)
Pulmonary Function Test	2844 (2.63)
Intubation	2420 (2.24)
Spirometry	1091 (1.01)
Bilevel positive airway pressure	710 (0.66)
Acute coronary syndrome	11008 (10.19)
Pneumonia	15386 (14.25)
Steroids	23249 (21.53)
Antihypertensives	7740 (7.17)
Short-acting bronchodilator	13088 (12.12)
Antihistiminic	17768 (16.45)
Respiratory Clearance	2791 (2.58)
Upper Respiratory Infection	1242 (1.15)
Antiarrythmic order	7650 (7.08)
Inhaled bronchodilators	122 (0.11)
Inhaled corticosteroid	78 (0.07)
Long-acting bronchodilator	91 (0.08)
Combination of inhaled bronchodilators	8 (0.001)
History of current emergency department visit	
Pneumonia	3503 (3.24)
Short-acting bronchodilator	5982 (5.54)
Steroids	4518 (4.18)
Antihypertensives	694 (0.64)
Acute coronary syndrome	2386 (2.21)
Antiarrthymic	1665 (1.54)
Antihistaminic	2624 (2.43)
Inhaled corticosteroid	146 (0.14)
Inhaled bronchodilators	304 (0.28)
Long-acting bronchodilator	420 (0.39)
Asthma	142 (0.13)
Upper Respiratory Infection	238 (0.22)
Respiratory Clearance	131 (0.12)
Combination of inhaled bronchodilators	3 (0.003)

C Type I and II Regret bounds

C.1 Notation and assumptions

We suppose there are n observations at time points $t = 1, ..., T$ for some $T \geq 2$. Consider any sequence of revision parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_T\}$, where $\theta_t \in \mathbb{R}^d$ for all $t = 1, 2, \ldots, T$, with unique values at times $\boldsymbol{\tau} = \{\tau_1, \tau_2, ..., \tau_s\}$, where $\tau_1 = 1 < \tau_2 < \ldots < \tau_s \leq T$. In other words, $\{\theta_{\tau_1}, \theta_{\tau_2}, \ldots, \theta_{\tau_s}\}$ denotes the sequence of values that the sequence $\boldsymbol{\theta}$ shifted over. Henceforth, we use $|\boldsymbol{\tau}|$ (rather than s) to indicate the number of times in $\boldsymbol{\tau}$. For ease of notation, we use the convention $\tau_{|\tau|+1} := T + 1$. Note that the variable $\tau_{|\tau|+1}$ is not part of the sequence $\boldsymbol{\tau}$ and is used purely to simplify the notation. We use $\boldsymbol{\tau}_{\text{locked}} := \{\tau_1\} = \{1\}$ to denote the shift times in the edge case of "locked" sequences $\boldsymbol{\theta}$ that do not shift over time. Let $D^{(t)}$ denote all the data observed up to time t.

The cumulative negative log-likelihood when using Bayesian inference at each time point is

$$L_{\text{BF}} = -\sum_{t=1}^{T} \log p\left(y_{t,\cdot} \mid z_{t,\cdot}, D^{(t-1)}\right),$$

where $p\left(y_{t,\cdot} \mid z_{t,\cdot}, D^{(t-1)}\right)$ is the posterior distribution at time t. The cumulative negative log-likelihoods for MarBLR and BLR are denoted by L_{MarBLR} and L_{BLR}, respectively, and are special cases of L_{BF} for their specific choice of priors. The MarBLR prior p_0 over $\boldsymbol{\theta}$ is defined using a Gaussian random walk with a homogeneous transition matrix as follows. Given $\theta_{\text{init}} \in \mathbb{R}^d$, $\Sigma_{\text{init}} \in \mathbb{R}^{d \times d}$, and some shift probability $\alpha \in [0, 1]$, let

$$\theta_1 \sim N(\theta_{\text{init}}, \Sigma_{\text{init}}), \quad W_1 = 1, \tag{8}$$

and for $t = 2, 3, \ldots, T$ let

$$\begin{aligned} \theta_t &= \theta_{t-1} + \beta_t W_t \\ W_t &\sim \text{Bernoulli}(\alpha) \\ \beta_t &\sim N(0, \delta^2 \Sigma_{\text{init}}). \end{aligned} \tag{9}$$

Note that $\boldsymbol{\tau} = \left\{\tau_1, \tau_2, \ldots, \tau_{|\boldsymbol{\tau}|}\right\}$ can be regarded as the indices at which the sequence $\{W_1, W_2, \ldots, W_T\}$ is 1-valued. In particular, having $\boldsymbol{\tau} = \boldsymbol{\tau}_{\text{locked}}$ implies that $W_1 = 1$ and $W_t = 0$ for all $t > 1$. The BLR prior is a special case where $\delta^2 = \alpha = 0$.

Type I regret compares BLR and MarBLR to locking the original revision parameters at its initial value θ_{init}, i.e. $\theta_t = \theta_{\text{init}}$ for all $t \in \{1, 2, \ldots, T\}$. The cumulative negative log-likelihood of the locked initial model is given by

$$L_{\text{locked}} = -\sum_{t=1}^{T} \log p\left(y_{t,\cdot} \mid z_{t,\cdot}; \theta_{\text{init}}\right).$$

Type II $\boldsymbol{\tau}$-regret compares BLR and MarBLR to the best sequence of parameters in retrospect for update times $\boldsymbol{\tau}$, denoted $\tilde{\theta}_{\tau_j}$ for $j = 1, ..., |\tau|$. Its cumulative negative log-likelihood is defined as

$$L_{\text{Dyn},\boldsymbol{\tau}} = -\sum_{t=1}^{T} \log p\left(y_{t,\cdot} \mid z_{t,\cdot}; \tilde{\theta}_t\right)$$

where $\tilde{\theta}_{\tau_j}$ for $j = 1, ..., |\tau|$ satisfy

$$\nabla \sum_{t=\tau_j}^{\tau_{j+1}-1} \log p\left(y_{t,\cdot} \mid z_{t,\cdot}; \theta\right)\Bigg|_{\theta=\tilde{\theta}_{\tau_j}} = 0, \quad \forall j = 1, \ldots, |\boldsymbol{\tau}|. \tag{10}$$

In addition, we introduce the notion of a distribution over the sequences $\boldsymbol{\theta}$. For such a distribution Q, its expected negative log-likelihood is given by

$$L_Q = E_Q\left[-\sum_{t=1}^{T} \log p\left(y_{t,\cdot} \mid z_{t,\cdot}; \theta_t\right)\right].$$

Given mean and variance parameters $\boldsymbol{\mu} = (\mu_t)_{t\in\boldsymbol{\tau}}$ and $\boldsymbol{\Sigma} = (\Sigma_t)_{t\in\boldsymbol{\tau}}$, we define $Q_{\boldsymbol{\tau},\boldsymbol{\mu},\Sigma}$ to be the distribution over $\boldsymbol{\theta}$ with shift times $\boldsymbol{\tau}$ where θ_{τ_j} for $j = 1, ..., |\boldsymbol{\tau}|$ are jointly independent and normally distributed per

$$\theta_{\tau_j} \sim N\left(\mu_{\tau_j}, \Sigma_j\right). \tag{11}$$

Some results in the following sections rely on the assumption that there exists a constant $c > 0$ such that

$$\left|\frac{\partial^2}{\partial w^2}\log p\left(y|z^\top\theta = w\right)\right| \leq c, \tag{12}$$

for all y and w. This always holds for logistic regression with $c \leq 1$.

C.2 Useful Results

Consider the prior distribution $p_0(\boldsymbol{\theta})$ over sequences $\boldsymbol{\theta}$. Let $p_0(\boldsymbol{\tau})$ be its marginal distribution over shift times $\boldsymbol{\tau}$ and $p_0(\boldsymbol{\theta} \mid \boldsymbol{\tau})$ be the conditional distribution over sequences $\boldsymbol{\theta}$ with shift times $\boldsymbol{\tau}$.

Lemma 1 (Variational bound)**.** *Consider any prior distribution p_0 over sequences $\boldsymbol{\theta}$. Given any $\boldsymbol{\tau}$ and any distribution Q, it holds that*

$$L_{\text{BF}} - L_Q \leq \mathrm{E}_{\boldsymbol{\tau}\sim Q}\left[\mathrm{KL}\left(Q(\boldsymbol{\theta}|\boldsymbol{\tau}) \mid\mid p_0(\boldsymbol{\theta} \mid \boldsymbol{\tau})\right)\right] + \mathrm{KL}\left(Q(\boldsymbol{\tau}) \mid\mid p_0(\boldsymbol{\tau})\right).$$

Proof. First, we can reexpress the cumulative negative log-likelihood of the Bayesian dynamical model by chaining the conditional probabilities as follows:

$$\begin{aligned} L_{\text{BF}} &= -\sum_{t=1}^{T}\log p\left(y_{t,\cdot} \mid z_{t,\cdot}, D^{(t-1)}\right) \\ &= -\log p\left((y_{t,\cdot})_{t=1,\ldots,T} \mid (z_{t,\cdot})_{t=1,\ldots,T}\right). \end{aligned}$$

Similarly, the cumulative negative log-likelihood of any sequence of calibration parameters can be written as

$$-\sum_{t=1}^{T}\log p\left(y_{t,\cdot} \mid z_{t,\cdot}; \theta_t\right) = -\log p\left((y_{t,\cdot})_{t=1,\ldots,T} \mid (z_{t,\cdot})_{t=1,\ldots,T}; \boldsymbol{\theta}\right).$$

Thus, the difference in the cumulative negative log-likelihood between the Bayesian dynamical model and any sequence of parameters is given by

$$L_{\text{BF}} - L_Q = \mathrm{E}_Q\left[\log\frac{p\left((y_{t,\cdot})_{t=1,\ldots,T} \mid (z_{t,\cdot})_{t=1,\ldots,T}; \boldsymbol{\theta}\right)}{p\left((y_{t,\cdot})_{t=1,\ldots,T} \mid (z_{t,\cdot})_{t=1,\ldots,T}\right)}\right].$$

By Bayes' Rule, the posterior distribution p_T over $\boldsymbol{\theta}$ with respect to the Bayesian dynamical model satisfies

$$p_T(\boldsymbol{\theta}) = \frac{p\left((y_{t,\cdot})_{t=1,\ldots,T} \mid (z_{t,\cdot})_{t=1,\ldots,T}; \boldsymbol{\theta}\right)p_0(\boldsymbol{\theta})}{p\left((y_{t,\cdot})_{t=1,\ldots,T} \mid (z_{t,\cdot})_{t=1,\ldots,T}\right)}.$$

Thus, we have that

$$\begin{aligned} L_{\text{BF}} - L_Q &= \mathrm{E}_Q\left[\log\frac{p_T(\boldsymbol{\theta})}{p_0(\boldsymbol{\theta})}\right] \\ &= \mathrm{E}_{\boldsymbol{\tau}\sim Q}\left[\mathrm{E}_{\boldsymbol{\theta}\sim Q(\cdot|\boldsymbol{\tau})}\left[\log\frac{p_T(\boldsymbol{\theta} \mid \boldsymbol{\tau})}{p_0(\boldsymbol{\theta} \mid \boldsymbol{\tau})}\right]\right] + \mathrm{E}_{\boldsymbol{\tau}\sim Q}\left[\log\frac{p_T(\boldsymbol{\tau})}{p_0(\boldsymbol{\tau})}\right]. \end{aligned} \tag{13}$$

Moreover, because the KL divergence is always positive, it holds that

$$\begin{aligned} \mathrm{E}_{\boldsymbol{\theta}\sim Q(\cdot|\boldsymbol{\tau})}\left[\log\frac{p_T(\boldsymbol{\theta} \mid \boldsymbol{\tau})}{p_0(\boldsymbol{\theta} \mid \boldsymbol{\tau})}\right] &= \mathrm{KL}\left(Q(\cdot|\boldsymbol{\tau}) \mid\mid p_0(\boldsymbol{\theta} \mid \boldsymbol{\tau})\right) - \mathrm{KL}\left(Q(\cdot|\boldsymbol{\tau}) \mid\mid p_T(\boldsymbol{\theta} \mid \boldsymbol{\tau})\right) \\ &\leq \mathrm{KL}\left(Q(\cdot|\boldsymbol{\tau}) \mid\mid p_0(\boldsymbol{\theta} \mid \boldsymbol{\tau})\right). \end{aligned} \tag{14}$$

Likewise,

$$E_{\boldsymbol{\tau}\sim Q}\left[\log\frac{p_T(\boldsymbol{\tau})}{p_0(\boldsymbol{\tau})}\right] \leq \mathrm{KL}\left(Q(\boldsymbol{\tau}) \mid\mid p_0(\boldsymbol{\tau})\right). \tag{15}$$

Finally, by combining equations (13) and (14) we arrive at the conclusion of this theorem. □

C.3 Type I regret results for MarBLR

Let the distribution p_0 be the MarBLR prior as defined per (8) and (9). For a given $\boldsymbol{\tau}$, let $Q^{(W)}_{\boldsymbol{\tau},\boldsymbol{\beta},\nu}$ be a Gaussian random walk with expected shifts β_j at known shift times τ_j for $j = 1, ..., |\tau|$. That is,

$$\theta_{\tau_j} - \theta_{\tau_{j-1}} \sim N\left(\beta_j, \nu^2 \Sigma_{\text{init}}\right). \tag{16}$$

and

$$\theta_{\tau_1} \sim N\left(\mu_1, \epsilon_1^2 \Sigma_{\text{init}}\right). \tag{17}$$

We begin with simplifying the KL divergence term in Lemma 1.

Lemma 2. *For any $\boldsymbol{\tau}$, consider the Gaussian random walk $Q^{(W)}_{\boldsymbol{\tau},\boldsymbol{\beta},\nu}$. We have that*

$$\begin{aligned}\mathrm{KL}\left(Q^{(W)}_{\boldsymbol{\tau},\boldsymbol{\beta},\nu} \,||\, p_0\left(\boldsymbol{\theta} \mid \boldsymbol{\tau}\right)\right) =& \frac{1}{2}\epsilon_1^2 d + \frac{1}{2}\left(\mu_1 - \theta_{\text{init}}\right)^\top \Sigma_{\text{init}}^{-1}\left(\mu_1 - \theta_{\text{init}}\right) - \frac{d|\tau|}{2} \\ &+ d(|\tau| - 1)\log\frac{\delta}{\nu} - d\log\epsilon_1 \\ &+ \frac{d\nu^2(|\tau|-1)}{2\delta^2} + \frac{1}{2\delta^2}\sum_{j=2}^{|\tau|}\beta_j^\top \Sigma_{\text{init}}^{-1}\beta_j.\end{aligned} \tag{18}$$

Proof. For ease of notation, let Θ_J be the space over sequences $(\theta_1, ..., \theta_J)$. Given the known times $\boldsymbol{\tau}$, there is a one-to-one mapping from sequences in $\Theta_{|\boldsymbol{\tau}|}$ to sequences in Θ_T with unique values at times $\boldsymbol{\tau}$. Let Q^{sub} be the probability distribution over $\Theta_{|\boldsymbol{\tau}|}$ as defined by $Q^{(W)}_{\boldsymbol{\tau},\boldsymbol{\beta},\nu}$. Likewise, let p_0^{sub} be the PDF over $\Theta_{|\boldsymbol{\tau}|}$ as defined by the conditional prior distribution $p_0\left(\cdot \mid \boldsymbol{\tau}\right)$.

We have that

$$\begin{aligned}&\mathrm{KL}\left(Q^{(W)}_{\boldsymbol{\tau},\boldsymbol{\beta},\nu} \,||\, p_0\left(\boldsymbol{\theta} \mid \boldsymbol{\tau}\right)\right) \\ =& \int \cdots \int Q^{\text{sub}}(\boldsymbol{\theta}) \log\frac{Q^{\text{sub}}(\boldsymbol{\theta})}{p_0^{\text{sub}}(\boldsymbol{\theta})} \mathrm{d}\boldsymbol{\theta} \\ =& \int Q^{\text{sub}}(\theta_1)\log\frac{Q^{\text{sub}}(\theta_1)}{p_0^{\text{sub}}(\theta_1)}\mathrm{d}\theta_1 + \sum_{j=1}^{|\tau|-1}\int\int Q^{\text{sub}}(\theta_j, \theta_{j+1})\log\frac{Q^{\text{sub}}(\theta_{j+1} \mid \theta_j)}{p_0^{\text{sub}}(\theta_{j+1} \mid \theta_j)}\mathrm{d}\theta_{j+1}\mathrm{d}\theta_j\end{aligned} \tag{19}$$

The first term in (19) is the KL divergence of two multivariate Normal distributions, $N(\mu_1, \epsilon_1^2\Sigma_{\text{init}})$ and $N(\theta_{\text{init}}, \Sigma_{\text{init}})$, and can be shown to be equal to

$$\frac{1}{2}\epsilon_1^2 d + \frac{1}{2}\left(\mu_1 - \theta_{\text{init}}\right)^\top \Sigma_{\text{init}}^{-1}\left(\mu_1 - \theta_{\text{init}}\right) - d\log\epsilon_1 - \frac{d}{2}. \tag{20}$$

Also, for $j = 1, ..., |\tau| - 1$, we have that each summand in the second term in (19) is equal to

$$\begin{aligned}&\int\int Q^{\text{sub}}(\theta_j, \theta_{j+1})\log\frac{Q^{\text{sub}}(\theta_{j+1} \mid \theta_j)}{p_0^{\text{sub}}(\theta_{j+1} \mid \theta_j)}\mathrm{d}\theta_{j+1}\mathrm{d}\theta_j \\ =& d\log\frac{\delta}{\nu} + \frac{1}{2\delta^2}\int\int Q^{\text{sub}}(\theta_j, \theta_{j+1})\left(\theta_{j+1} - \theta_j\right)^\top \Sigma_{\text{init}}^{-1}\left(\theta_{j+1} - \theta_j\right)\mathrm{d}\theta_{j+1}\mathrm{d}\theta_j - \frac{d}{2}.\end{aligned} \tag{21}$$

By the definition of $Q^{(W)}_{\boldsymbol{\tau},\boldsymbol{\beta},\nu}$, we have that $\Delta_{j+1} = \theta_{\tau_{j+1}} - \theta_{\tau_j} \sim N\left(\beta_{j+1}, \nu^2\Sigma_{\text{init}}\right)$. Thus,

$$\int Q^{\text{sub}}(\theta_j, \theta_{j+1})\left(\theta_{j+1} - \theta_j\right)^\top \Sigma_{\text{init}}^{-1}\left(\theta_{j+1} - \theta_j\right)\mathrm{d}\theta_{j+1}\mathrm{d}\theta_j = d\nu^2 + \beta_{j+1}^\top\Sigma_{\text{init}}^{-1}\beta_{j+1}.$$

Plugging in the above result into (21), we have that

$$\int\int Q^{\text{sub}}(\theta_j, \theta_{j+1})\log\frac{Q^{\text{sub}}(\theta_{j+1} \mid \theta_j)}{p_0^{\text{sub}}(\theta_{j+1} \mid \theta_j)}\mathrm{d}\theta_{j+1}\mathrm{d}\theta_j = d\log\frac{\delta}{\nu} + \frac{1}{2\delta^2}\left(d\nu^2 + \beta_{j+1}^\top\Sigma_{\text{init}}^{-1}\beta_{j+1}\right) - \frac{d}{2}. \tag{22}$$

Combining the results (19), (20) and (22), we attain the desired conclusion. ☐

To bound the Type I regret for MarBLR, we compare the regret via the intermediary Q with marginal distribution over $\boldsymbol{\tau}$ the same as p_0 and the conditional distribution given τ to be $Q^{(W)}_{\boldsymbol{\tau},\boldsymbol{\beta},\nu}$ with $\boldsymbol{\beta}_j = 0$ for all $j = 2, ..., |\tau|$. That is, the regret is decomposed into

$$(L_{\text{BF}} - L_Q) + (L_Q - L_{\text{locked}}). \tag{23}$$

We bound $L_{\text{BF}} - L_Q$ by marginalizing Lemma 2 over $\boldsymbol{\tau}$ as follows.

Lemma 3. *Let the distribution p_0 be defined as above. Let distribution Q over $\boldsymbol{\theta}$ have the same distribution over $\boldsymbol{\tau}$ as p_0, with θ_1 distributed $N(\theta_{\text{init}}, \epsilon_1^2 \Sigma_{\text{init}})$, and $Q(\cdot|\boldsymbol{\tau})$ be a zero-centered Gaussian random walk $\boldsymbol{\beta}_j = 0$ for all $j = 2, ..., |\tau|$. Let $\xi = \mathrm{E}_{p_0} |\boldsymbol{\tau}|$. We have that*

$$L_{\text{BF}} - L_Q \leq \frac{1}{2}\epsilon_1^2 d - \frac{d\xi}{2} + d(\xi - 1)\log\frac{\delta}{\nu} - d\log\epsilon_1 + \frac{d\nu^2(\xi - 1)}{2\delta^2}. \tag{24}$$

Proof. Taking the expectation of (18) from Lemma 2 with respect to $\boldsymbol{\tau}$ under the additional assumptions of this Lemma, and plugging the result into Lemma 1 yields the desired conclusion. □

Next we bound $L_Q - L_{\text{locked}}$.

Lemma 4. *Assume that there is a $c > 0$ that bounds the second derivative as in (12). Assume that there is an R such that $\frac{1}{n(\tau_{j+1} - \tau_j)} \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^n z_{t,i} z_{t,i}^\top \preceq R^2 I$ for all $j \in \{1, 2, \ldots, |\boldsymbol{\tau}|\}$. Let $Q^{(W)}_{\boldsymbol{\tau},\boldsymbol{\beta},\nu}$ be the zero-centered Gaussian random walk with $\mu_1 = \theta_{\text{init}}$. Then it holds that*

$$L_{Q^{(W)}_{\boldsymbol{\tau},\boldsymbol{\beta},\nu}} - L_{\text{locked}} \leq \frac{cnR^2}{2} \operatorname{Tr}(\Sigma_{\text{init}}) \left(T\epsilon_1^2 + \nu^2 \sum_{j=2}^{|\tau|} (\tau_{j+1} - \tau_j)(j - 1) \right).$$

Proof. We use a Taylor expansion. For $j = 1, ..., |\tau|$, there is some θ_{mid} such that

$$\begin{aligned} -\sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^n \log p\left(y_{t,i} \mid z_{t,i}; \theta_{\tau_j}\right) = &- \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^n \log p\left(y_{t,i} \mid z_{t,i}; \theta_{\text{init}}\right) - \nabla_\theta \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^n \log p\left(y_{t,i} \mid z_{t,i}; \theta\right)\Bigg|_{\theta=\theta_{\text{init}}}^\top \left(\theta_{\tau_j} - \theta_{\text{init}}\right) \\ &- \frac{1}{2}\left(\theta_{\tau_j} - \theta_{\text{init}}\right)^\top \nabla_\theta^2 \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^n \log p\left(y_{t,i} \mid z_{t,i}; \theta\right)\Bigg|_{\theta=\theta_{mid}} \left(\theta_{\tau_j} - \theta_{\text{init}}\right). \end{aligned} \tag{25}$$

Note that

$$\left(\theta_{\tau_j} - \theta_{\text{init}}\right)^\top \nabla_\theta^2 \log p\left(y \mid z; \theta\right) \left(\theta_{\tau_j} - \theta_{\text{init}}\right) = \frac{\partial^2}{\partial w^2} \log p\left(y|w\right) \left(z^\top \left(\theta_{\tau_j} - \theta_{\text{init}}\right)\right)^2,$$

where $w = z^\top \theta$ is the predicted logit. Using equation (12) it follows that

$$\left| \frac{1}{2}\left(\theta_{\tau_j} - \theta_{\text{init}}\right)^\top \nabla_\theta^2 \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^n \log p\left(y_{t,i} \mid z_{t,i}; \theta\right)\Bigg|_{\theta=\theta_{mid}} \left(\theta_{\tau_j} - \theta_{\text{init}}\right) \right| \tag{26}$$

$$\leq \frac{c}{2} \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^n \left(z_{t,i}^\top \left(\theta_{\tau_j} - \theta_{\text{init}}\right)\right)^2 \tag{27}$$

$$= \frac{c}{2}\left(\theta_{\tau_j} - \theta_{\text{init}}\right)^\top \left(\sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^n z_{t,i} z_{t,i}^\top \right) \left(\theta_{\tau_j} - \theta_{\text{init}}\right). \tag{28}$$

Because the expected value of θ with respect to Q is θ_{init}, we have the following after taking the expectation of equation (25) combined with equation (26):

$$
\begin{aligned}
L_Q &= \mathrm{E}_Q\left[-\sum_{t=1}^{T}\sum_{i=1}^{n}\log p\left(y_{t,i} \mid z_{t,i};\theta_t\right)\right] \\
&\le L_{\text{locked}} + \sum_{j=1}^{|\tau|}\frac{c}{2}\,\mathrm{E}_Q\left[\left(\theta_{\tau_j}-\theta_{\text{init}}\right)^\top\left(\sum_{t=\tau_j}^{\tau_{j+1}-1}\sum_{i=1}^{n} z_{t,i}z_{t,i}^\top\right)\left(\theta_{\tau_j}-\theta_{\text{init}}\right)\right].
\end{aligned}
$$

Assuming there exists some R^2 that satisfies the lemma assumptions, the following holds after taking the expectation with respect to Q:

$$
\begin{aligned}
\mathrm{E}_Q\left[\left(\theta_{\tau_j}-\theta_{\text{init}}\right)^\top\left(\sum_{t=\tau_j}^{\tau_{j+1}-1}\sum_{i=1}^{n} z_{t,i}z_{t,i}^\top\right)\left(\theta_{\tau_j}-\theta_{\text{init}}\right)\right] &\le (\tau_{j+1}-\tau_j)nR^2\,\mathrm{E}_Q\left\|\theta_{\tau_j}-\theta_{\text{init}}\right\|^2 \\
&= (\tau_{j+1}-\tau_j)nR^2\left(\epsilon_1^2+(j-1)\nu^2\right)\mathrm{Tr}(\Sigma_{\text{init}}).
\end{aligned}
$$

After summing over $j = 1, ..., |\tau|$, we reach our desired result. □

We combine the two prior lemmas to obtain the following bound on the Type I error for MarBLR.

Theorem 5 (Type I regret for MarBLR)**.** *Let $\xi = \mathrm{E}_{p_0}|\boldsymbol{\tau}|$ denote the expected number of shift times be denoted. The Type I regret for MarBLR is bounded as follows:*

$$
\begin{aligned}
L_{\text{MarBLR}} - L_{\text{locked}} \le \frac{d}{2}\log\left(1+\frac{cnR^2T\,\mathrm{Tr}(\Sigma_{\text{init}})}{d}\right) \\
+ \frac{d\alpha(T-1)}{2}\log\left(1+\frac{\delta^2 cnR^2T\,\mathrm{Tr}(\Sigma_{\text{init}})}{2d}\right).
\end{aligned}
$$

Proof. First, note that under the MarBLR prior p_0 over shift times $\boldsymbol{\tau}$ as defined previously, we have that

$$
\mathrm{E}_{p_0}\left[\sum_{j=2}^{|\tau|}(\tau_{j+1}-\tau_j)(j-1)\right] = \mathrm{E}_{p_0}\left[\sum_{t=2}^{T} W_t(T+1-t)\right] = \frac{\alpha}{2}T(T-1),
$$

and $\xi = \mathrm{E}_{p_0}|\boldsymbol{\tau}| = \alpha(T-1)+1$.

Thus, summing the upper bounds from Lemmas 3 and 4 and taking expectations with respect to $\boldsymbol{\tau} \sim p_0$, we have that

$$
\begin{aligned}
L_{\text{BF}} - L_{\text{locked}} &\le \frac{1}{2}\epsilon_1^2 d - \frac{d\alpha(T-1)}{2} - \frac{d}{2} + d\alpha(T-1)\log\frac{\delta}{\nu} - d\log\epsilon_1 + \frac{d\alpha(T-1)}{2\delta^2}\nu^2 \qquad (29)\\
&\quad + \frac{cnR^2T}{2}\mathrm{Tr}(\Sigma_{\text{init}})\left(\epsilon_1^2+\frac{\alpha}{2}(T-1)\nu^2\right). \qquad (30)
\end{aligned}
$$

We minimize the upper bound by selecting

$$
\epsilon_1^2 = \frac{d}{d+cnR^2T\,\mathrm{Tr}(\Sigma_{\text{init}})}
$$

and

$$
\nu^2 = \frac{d}{\frac{d}{\delta^2}+\frac{c}{2}nR^2T\,\mathrm{Tr}(\Sigma_{\text{init}})}
$$

to obtain the upper bound

$$
\frac{d}{2}\log\left(1+\frac{cnR^2T\,\mathrm{Tr}(\Sigma_{\text{init}})}{d}\right) + \frac{d\alpha(T-1)}{2}\log\left(1+\frac{\delta^2 cnR^2T\,\mathrm{Tr}(\Sigma_{\text{init}})}{2d}\right).
$$

□

C.4 Type II $\boldsymbol{\tau}$-regret results for BLR

Let $\tilde{\theta}_{\tau_{\text{locked}}}$ be the minimizer of the cumulative log-likelihood of the locked model, i.e., $\tilde{\theta}_{\tau_{\text{locked}}}$ satisfies that

$$\nabla \sum_{t=1}^{T} \sum_{i=1}^{n} \log p\left(y_{t,i} \mid z_{t,i}; \theta\right) \Big|_{\theta=\tilde{\theta}_{\tau_{\text{locked}}}} = 0.$$

Let $\tilde{Q}$ denote the distribution $Q_{\boldsymbol{\tau}_{\text{locked}}, \tilde{\theta}_{\tau_{\text{locked}}}, \epsilon^2 \Sigma_{\text{init}}}$ (defined according to section C.1 and equation (11) with the parameters specified here). That is, we have that

$$\tilde{Q}(\theta_1) = Q_{\boldsymbol{\tau}_{\text{locked}}, \tilde{\theta}_{\tau_{\text{locked}}}, \epsilon^2 \Sigma_{\text{init}}}(\theta_1) = N\left(\tilde{\theta}_{\tau_{locked}}, \epsilon^2 \Sigma_{\text{init}}\right),$$

and $\theta_t = \theta_1$ for all $t \in \{2, 3, \ldots, T\}$.

We bound the difference in the cumulative negative log-likelihood, $L_{\text{BLR}} - L_{\text{Dyn},\boldsymbol{\tau}}$, by breaking it into two summands

$$L_{\text{BLR}} - L_{\text{Dyn},\boldsymbol{\tau}} = \left(L_{\text{BLR}} - L_{\tilde{Q}}\right) + \left(L_{\tilde{Q}} - L_{\text{Dyn},\boldsymbol{\tau}}\right). \tag{31}$$

We have already bounded the first summand by Lemmas 1 and 2. We just need to bound the second summand.

Lemma 6. *Assume that the second derivative is bounded by a constant c as shown in equation (12), and that there are $R_{\tau_1}, R_{\tau_2}, \ldots, R_{\tau_{|\boldsymbol{\tau}|}}$ such that*

$$\frac{1}{n(\tau_{j+1} - \tau_j)} \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} z_{t,i} z_{t,i}^\top \preceq R_j^2 I.$$

It holds that

$$L_{\tilde{Q}} - L_{\text{Dyn},\boldsymbol{\tau}} \leq \frac{cn \sum_{j=1}^{|\boldsymbol{\tau}|} R_j^2 \left(\tau_{j+1} - \tau_j\right)}{2} \epsilon^2 \operatorname{Tr}(\Sigma_{\text{init}}) + \frac{cn}{2} \sum_{j=1}^{|\boldsymbol{\tau}|} R_j^2 \left(\tau_{j+1} - \tau_j\right) \left\| \tilde{\theta}_{\tau_{locked}} - \tilde{\theta}_{\tau_j} \right\|^2.$$

Proof. Because $\tilde{\theta}_{\tau_j}$ is the minimizer of $\nabla_\theta \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} \log p\left(y_{t,i} \mid z_{t,i}; \theta\right)$, per Taylor's expansion there is some θ_{mid} such that

$$\begin{aligned} - \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} \log p\left(y_{t,i} \mid z_{t,i}; \theta\right) = &- \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} \log p\left(y_{t,i} \mid z_{t,i}; \tilde{\theta}_{\tau_j}\right) \\ &- \frac{1}{2} \left(\theta - \tilde{\theta}_{\tau_j}\right)^\top \nabla_\theta^2 \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} \log p\left(y_{t,i} \mid z_{t,i}; \theta\right) \Big|_{\theta=\theta_{\text{mid}}} \left(\theta - \tilde{\theta}_{\tau_j}\right). \end{aligned}$$

Following the same arguments as in the proof of Lemma 4, we have that

$$\begin{aligned} \mathrm{E}_{\tilde{Q}} \left[- \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} \log p\left(y_{t,i} \mid z_{t,i}; \theta\right) \right] \leq &- \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} \log p\left(y_{t,i} \mid z_{t,i}; \tilde{\theta}_{\tau_j}\right) \\ &+ \frac{c}{2} \mathrm{E}_{\tilde{Q}} \left[\left(\theta_1 - \tilde{\theta}_{\tau_j}\right)^\top \left(\sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} z_{t,i} z_{t,i}^\top \right) \left(\theta_1 - \tilde{\theta}_{\tau_j}\right) \right]. \end{aligned}$$

Taking expectation with respect to $\tilde{Q}$, we note that

$$
\begin{aligned}
&\mathrm{E}_{\tilde{Q}}\left[\left(\theta_1-\tilde{\theta}_{\tau_j}\right)^\top\left(\sum_{t=\tau_j}^{\tau_{j+1}-1}\sum_{i=1}^{n} z_{t,i}z_{t,i}^\top\right)\left(\theta_1-\tilde{\theta}_{\tau_j}\right)\right]\\
&=\mathrm{E}_{\tilde{Q}}\left[\left(\theta_1-\tilde{\theta}_{\tau_{\text{locked}}}\right)^\top\left(\sum_{t=\tau_j}^{\tau_{j+1}-1}\sum_{i=1}^{n} z_{t,i}z_{t,i}^\top\right)\left(\theta_1-\tilde{\theta}_{\tau_{\text{locked}}}\right)\right]+\left(\tilde{\theta}_{\tau_{\text{locked}}}-\tilde{\theta}_{\tau_j}\right)^\top\left(\sum_{t=\tau_j}^{\tau_{j+1}-1}\sum_{i=1}^{n} z_{t,i}z_{t,i}^\top\right)\left(\tilde{\theta}_{\tau_{\text{locked}}}-\tilde{\theta}_{\tau_j}\right)\\
&\leq\left(\tau_{j+1}-\tau_j\right)nR_j^2\epsilon^2\,\mathrm{Tr}\left(\Sigma_{\text{init}}\right)+\left(\tau_{j+1}-\tau_j\right)nR_j^2\left\|\tilde{\theta}_{\tau_{\text{locked}}}-\tilde{\theta}_{\tau_j}\right\|^2.
\end{aligned}
$$

We arrive at our results after summing over all $j=1,...,|\tau|$. ☐

Theorem 7 (Type II regret for BLR)**.** *Assume that there is an R such that $\frac{1}{n(\tau_{j+1}-\tau_j)}\sum_{t=\tau_j}^{\tau_{j+1}-1}\sum_{i=1}^{n} z_{t,i}z_{t,i}^\top\preceq R^2I$ for all $j\in\{1,2,\ldots,|\boldsymbol{\tau}|\}$. It holds that*

$$
\begin{aligned}
L_{\text{BLR}}-L_{\text{Dyn},\boldsymbol{\tau}}\leq{}&\frac{1}{2}\left(\tilde{\theta}_{\tau_{\text{locked}}}-\theta_{\text{init}}\right)^\top\Sigma_{\text{init}}^{-1}\left(\tilde{\theta}_{\tau_{\text{locked}}}-\theta_{\text{init}}\right)+\frac{d}{2}\log\left(\frac{d+cnTR^2\,\mathrm{Tr}(\Sigma_{\text{init}})}{d}\right)\\
&+\frac{cnR^2}{2}\sum_{j=1}^{|\tau|}\left(\tau_{j+1}-\tau_j\right)\left\|\tilde{\theta}_{\tau_{\text{locked}}}-\tilde{\theta}_{\tau_j}\right\|^2.
\end{aligned}
$$

Proof. To bound the first summand of decomposition (31), we use Lemmas 1 and 2 and the fact that $p_0(\boldsymbol{\tau}_{\text{locked}})=1$ under BLR. We use Lemma 6 to bound the second summand of decomposition (31). Thus, we obtain

$$
\begin{aligned}
L_{\text{BLR}}-L_{\text{locked}}\leq{}&\frac{1}{2}\epsilon^2d+\frac{1}{2}\left(\tilde{\theta}_{\tau_{locked}}-\theta_{\text{init}}\right)^\top\Sigma^{-1}\left(\tilde{\theta}_{\tau_{locked}}-\theta_{\text{init}}\right)+\frac{d}{2}-d\log(\epsilon)+\frac{cnTR^2}{2}\epsilon^2\,\mathrm{Tr}(\Sigma_{\text{init}})\\
&+\frac{cnR}{2}\sum_{j=1}^{|\tau|}\left(\tau_{j+1}-\tau_j\right)\left\|\tilde{\theta}_{\tau_{locked}}-\tilde{\theta}_{\tau,j}\right\|^2.
\end{aligned}
$$

Choosing $\epsilon^2=\frac{d}{d+cnTR^2\,\mathrm{Tr}(\Sigma_{\text{init}})}$ will minimize this expression, which yields the desired conclusion. ☐

C.5 Type II $\boldsymbol{\tau}$-regret results for MarBLR

As before, we bound the difference in the cumulative negative log-likelihood, $L_{\text{BF}}-L_{\text{Dyn},\boldsymbol{\tau}}$, by breaking it into two summands

$$
L_{\text{BF}}-L_{\text{Dyn},\boldsymbol{\tau}}=\left(L_{\text{BF}}-L_{Q_{\boldsymbol{\tau}',\tilde{\boldsymbol{\theta}}',\boldsymbol{\epsilon}^2\Sigma_{\text{init}}}}\right)+\left(L_{Q_{\boldsymbol{\tau}',\tilde{\boldsymbol{\theta}}',\boldsymbol{\epsilon}^2\Sigma_{\text{init}}}}-L_{\text{Dyn},\boldsymbol{\tau}}\right). \tag{32}
$$

Thus the proof proceeds by comparing against an intermediary distribution $Q_{\boldsymbol{\tau}',\tilde{\boldsymbol{\theta}}',\boldsymbol{\epsilon}^2\Sigma_{\text{init}}}$ defined per (11), where $\boldsymbol{\tau}'$ be any subsequence of $\boldsymbol{\tau}$ with $\tau_1'=1$, $\tilde{\boldsymbol{\theta}}':=(\tilde{\theta}_t)_{t\in\boldsymbol{\tau}'}$, and $\boldsymbol{\epsilon}^2=\left(\epsilon_1^2,\epsilon_2^2,\ldots,\epsilon_{|\boldsymbol{\tau}'|}^2\right)$. This intermediary distribution is centered around a dynamic oracle that may evolve slower than than the specified update times $\boldsymbol{\tau}$. The final Type II regret bound will depend on $\boldsymbol{\tau}'$. Optimizing our choice of $\boldsymbol{\tau}'$ can lead to tighter Type Ii regret bounds, particularly when α in the MarBLR prior is small and $|\tau|$ is large.

We use the following lemma to bound the first summand of (32).

Lemma 8. *Consider the distribution $Q_{\boldsymbol{\tau},\boldsymbol{\mu},\boldsymbol{\epsilon}^2\Sigma_{\text{init}}}$ as defined above, and the MarBLR prior p_0 as defined per (8) and (9). For any $\boldsymbol{\tau}$, $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}^2$, we have that*

$$
\begin{aligned}
&KL\left(Q_{\boldsymbol{\tau},\boldsymbol{\mu},\boldsymbol{\epsilon}^2\Sigma_{\text{init}}}\;||\;p_0\left(\boldsymbol{\theta}\mid\tau\right)\right)\\
&=\frac{d}{2}\epsilon_1^2+\frac{1}{2}\left(\mu_1-\theta_{\text{init}}\right)^\top\Sigma_{\text{init}}^{-1}\left(\mu_1-\theta_{\text{init}}\right)-d\log\epsilon_1-\frac{d}{2}|\boldsymbol{\tau}|+(|\boldsymbol{\tau}|-1)d\log\delta\\
&+\sum_{t=2}^{|\boldsymbol{\tau}|}\left[\frac{1}{2\delta^2}\left(d\left(\epsilon_{t-1}^2+\epsilon_t^2\right)+\left(\mu_t-\mu_{t-1}\right)^\top\Sigma_{\text{init}}^{-1}\left(\mu_t-\mu_{t-1}\right)\right)-d\log\epsilon_t\right].
\end{aligned}
$$

Proof. We define Θ_J and p_0^{sub} as in Lemma 2. We define Q^{sub} as the distribution over Θ_J as defined by $Q_{\boldsymbol{\tau},\boldsymbol{\mu},\boldsymbol{\epsilon}^2\Sigma_{\text{init}}}$. We have that

$$KL\left(Q_{\tau,(\mu_t)_{t\in\tau}} \mid\mid p_0\left(\boldsymbol{\theta} \mid \tau\right)\right) = KL\left(Q^{\text{sub}} \mid\mid p_0^{\text{sub}}\right)$$
$$= \int \cdots \int Q^{\text{sub}}(\boldsymbol{\theta}) \sum_{t=1}^{|\boldsymbol{\tau}|} \log \frac{Q^{\text{sub}}\left(\theta_t\right)}{p_0^{\text{sub}}\left(\theta_t \mid \theta_{t-1}\right)} d\theta_{|\boldsymbol{\tau}|} \cdots d\theta_1, \tag{33}$$

because θ_t in Q^{sub} are jointly independent and θ_t in p_0^{sub} only depend on θ_{t-1}. As such,

$$KL\left(Q^{\text{sub}} \mid\mid p_0^{\text{sub}}\right) = \int Q^{\text{sub}}\left(\theta_1\right) \log \frac{Q^{\text{sub}}\left(\theta_1\right)}{p_0^{\text{sub}}\left(\theta_1\right)} d\theta_1 \tag{34}$$
$$+ \sum_{t=2}^{|\boldsymbol{\tau}|} \int\int Q^{\text{sub}}\left(\theta_{t-1}, \theta_t\right) \log \frac{Q^{\text{sub}}\left(\theta_t\right)}{p_0^{\text{sub}}\left(\theta_t \mid \theta_{t-1}\right)} d\theta_t d\theta_{t-1}. \tag{35}$$

The first term (34) is the KL divergence of two multivariate Normal distributions, $N(\mu_1, \epsilon_1^2\Sigma_{\text{init}})$ and $N(\theta_{\text{init}}, \Sigma_{\text{init}})$, and can be shown to be equal to

$$\int Q^{\text{sub}}\left(\theta_1\right) \log \frac{Q^{\text{sub}}\left(\theta_1\right)}{p_0^{\text{sub}}\left(\theta_1\right)} d\theta_1 = \frac{1}{2}\epsilon_1^2 d + \frac{1}{2}\left(\mu_1 - \theta_{\text{init}}\right)^\top \Sigma_{\text{init}}^{-1}\left(\mu_1 - \theta_{\text{init}}\right) - d\log\epsilon_1 - \frac{d}{2}. \tag{36}$$

Next each term in the summation of (35) is equal to

$$\int\int Q^{\text{sub}}(\theta_{t-1}, \theta_t) \log \frac{Q^{\text{sub}}(\theta_t)}{p_0^{\text{sub}}(\theta_t \mid \theta_{t-1})} \mathrm{d}\theta_t \mathrm{d}\theta_{t-1}$$
$$= d\log\frac{\delta}{\epsilon_t} + \frac{1}{2\delta^2}\int\int Q^{\text{sub}}(\theta_{t-1}, \theta_t)\left(\theta_t - \theta_{t-1}\right)^\top \Sigma_{\text{init}}^{-1}\left(\theta_t - \theta_{t-1}\right) \mathrm{d}\theta_t \mathrm{d}\theta_{t-1} - \frac{d}{2}. \tag{37}$$

We note that under Q^{sub} it holds that

$$\left(\theta_t - \theta_{t-1}\right) \sim N\left(\mu_t - \mu_{t-1}, (\epsilon_{t-1}^2 + \epsilon_t^2)\Sigma_{\text{init}}\right).$$

Therefore, (37) simplifies to

$$\int\int Q^{\text{sub}}(\theta_{t-1}, \theta_t) \log \frac{Q^{\text{sub}}(\theta_t)}{p_0^{\text{sub}}(\theta_t \mid \theta_{t-1})} \mathrm{d}\theta_t \mathrm{d}\theta_{t-1}$$
$$= d\log\frac{\delta}{\epsilon_t} + \frac{1}{2\delta^2}\left(d\left(\epsilon_{t-1}^2 + \epsilon_t^2\right) + (\mu_t - \mu_{t-1})^\top \Sigma_{\text{init}}^{-1}(\mu_t - \mu_{t-1})\right) - \frac{d}{2}. \tag{38}$$

Plugging (36) and (38) into (34) and (35) gives us the desired result. ☐

Next we need to bound the second summand of (32).

Lemma 9. *Suppose there is a constant c that bounds the second derivative as in* (12)*. Assume that there is an R such that $\frac{1}{n(\tau_{j+1}-\tau_j)}\sum_{t=\tau_j}^{\tau_{j+1}-1}\sum_{i=1}^n z_{t,i}z_{t,i}^\top \preceq R^2 I$ for all $j \in \{1, 2, \ldots, |\boldsymbol{\tau}|\}$. Then it holds that*

$$L_{Q_{\boldsymbol{\tau}',\tilde{\boldsymbol{\theta}}',\boldsymbol{\epsilon}^2\Sigma_{\text{init}}}} - L_{\text{Dyn},\boldsymbol{\tau}} \leq \frac{1}{2}cnR^2\sum_{j=1}^{|\boldsymbol{\tau}|}(\tau_{j+1} - \tau_j)\left(\epsilon_{k(j)}^2 \operatorname{Tr}(\Sigma_{\text{init}}) + \left\|\tilde{\theta}_{\tau'_{k(j)}} - \tilde{\theta}_{\tau_j}\right\|^2\right)$$

where $k(j) := \max\{k : \tau'_k \leq \tau_j\}$.

Proof. For the ease of notation denote $\tilde{Q} := Q_{\boldsymbol{\tau}',\tilde{\boldsymbol{\theta}}',\boldsymbol{\epsilon}^2\Sigma_{\text{init}}}$. It holds that

$$L_{\tilde{Q}} - L_{\text{Dyn},\boldsymbol{\tau}}$$
$$= \sum_{j=1}^{|\boldsymbol{\tau}|}\left(E_{\tilde{Q}}\left[\sum_{t=\tau_j}^{\tau_{j+1}-1}\sum_{i=1}^{n} -\log p\left(y_{t,i} \mid z_{t,i}; \theta_t\right) + \log p\left(y_{t,i} \mid z_{t,i}; \tilde{\theta}_{\tau_j}\right)\right]\right)$$

Recall that for any sequence $\boldsymbol{\theta}$ drawn from $\tilde{Q}$, for any $j = 1, \ldots, |\boldsymbol{\tau}|$, the parameters θ_t are constant over $t = \tau'_j, \ldots, \tau'_{j+1} - 1$. Taking a Taylor expansion, there exists some θ_{mid} such that

$$
\begin{aligned}
-\sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} \log p\left(y_{t,i}|z_{t,i};\theta\right) = &-\sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} \log p\left(y_{t,i}\middle|z_{t,i};\tilde{\theta}_{\tau_j}\right) \\
&- \nabla_\theta \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} \log p\left(y_{t,i}|z_{t,i};\theta\right)\Bigg|_{\theta=\tilde{\theta}_{\tau_j}}^{\top} \left(\theta - \tilde{\theta}_{\tau_j}\right) \\
&- \frac{1}{2}\left(\theta - \tilde{\theta}_{\tau_j}\right)^{\top} \nabla_\theta^2 \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} \log p\left(y_{t,i}|z_{t,i};\theta\right)\Bigg|_{\theta=\theta_{mid}} \left(\theta - \tilde{\theta}_{\tau_j}\right).
\end{aligned} \tag{39}
$$

Since $\boldsymbol{\tau}'$ is a subsequence of $\boldsymbol{\tau}$, for $\boldsymbol{\theta} \sim \tilde{Q}$ we have that $\theta_t = \theta_{\tau'_{k(j)}}$ for all $t = \tau_j, ..., \tau_{j+1} - 1$, where $k(j) := \max\{k : \tau'_k \leq \tau_j\}$. Thus, we can use the above decomposition to evaluate (39) with $\theta_t = \theta_{\tau'_{k_j}}$ in place of θ.

By the definition of $\tilde{\theta}_{\tau_j}$, the gradient in the expression above is zero, so the second term is equal to zero. Because we assumed the second derivative was bounded by c as in (12), the expression simplifies to the bound

$$
\begin{aligned}
-\sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} \log p\left(y_{t,i} \mid z_{t,i};\theta_t\right) \leq &-\sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} \log p\left(y_{t,i} \mid z_{t,i};\tilde{\theta}_{\tau_j}\right) \\
&+ \frac{c}{2}\left(\theta_{\tau'_{k(j)}} - \tilde{\theta}_{\tau_j}\right)^{\top} \left(\sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} z_{t,i} z_{t,i}^{\top}\right) \left(\theta_{\tau'_{k(j)}} - \tilde{\theta}_{\tau_j}\right).
\end{aligned}
$$

Assuming there exists some R^2 that satisfies the lemma assumptions, it follows that

$$
\begin{aligned}
&\mathrm{E}_{\tilde{Q}}\left[\left(\theta_{\tau'_{k(j)}} - \tilde{\theta}_{\tau_j}\right)^{\top} \left(\sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} z_{t,i} z_{t,i}^{\top}\right) \left(\theta_{\tau'_{k(j)}} - \tilde{\theta}_{\tau_j}\right)\right] \\
&\leq (\tau_{j+1} - \tau_j) n R^2 \,\mathrm{E}_{\tilde{Q}} \left\|\theta_{\tau'_{k(j)}} - \tilde{\theta}_{\tau_j}\right\|^2 \\
&= (\tau_{j+1} - \tau_j) n R^2 \left(\epsilon_{k(j)}^2 \operatorname{Tr}(\Sigma_{\text{init}}) + \left\|\tilde{\theta}_{\tau'_{k(j)}} - \tilde{\theta}_{\tau_j}\right\|^2\right).
\end{aligned}
$$

We finish the proof by summing over j. □

We combine the results to get the following bound.

Theorem 10 (Type II regret for MarBLR)**.** *Suppose there is a constant* c *that bounds the second derivative as in* (12)*. Assume that there is an* R *such that* $\frac{1}{n(\tau_{j+1}-\tau_j)} \sum_{t=\tau_j}^{\tau_{j+1}-1} \sum_{i=1}^{n} z_{t,i} z_{t,i}^{\top} \preceq R^2 I$ *for all* $j \in \{1, 2, \ldots, |\boldsymbol{\tau}|\}$*. Let* $\boldsymbol{\tau}'$ *be any subsequence of the sequence of shift times* $\boldsymbol{\tau}$*. Then it holds that*

$$
\begin{aligned}
&L_{\text{MarBLR}} - L_{\text{Dyn},\boldsymbol{\tau}} \\
&\leq \frac{1}{2}\left(\tilde{\theta}_1 - \theta_{\text{init}}\right)^{\top} \Sigma_{\text{init}}^{-1} \left(\tilde{\theta}_1 - \theta_{\text{init}}\right) + \frac{d}{2} \log\left(1 + \frac{1}{\delta^2} + \frac{cnR^2 \operatorname{Tr}(\Sigma_{\text{init}})\left(\tau'_2 - \tau'_1\right)}{d}\right) \\
&+ \frac{1}{2} \sum_{t=2}^{|\boldsymbol{\tau}'|} \left[\frac{1}{\delta^2}\left(\tilde{\theta}_{\tau'_t} - \tilde{\theta}_{\tau'_{t-1}}\right)^{\top} \Sigma_{\text{init}}^{-1} \left(\tilde{\theta}_{\tau'_t} - \tilde{\theta}_{\tau'_{t-1}}\right) + d \log\left(\frac{2}{\delta^2} + \frac{cnR^2 \operatorname{Tr}(\Sigma_{\text{init}})\left(\tau'_{j+1} - \tau'_j\right)}{d}\right)\right] \\
&- \log p_0(\boldsymbol{\tau}') + (|\boldsymbol{\tau}'| - 1) d \log \delta + \frac{1}{2} cnR^2 \sum_{j=1}^{|\boldsymbol{\tau}|} (\tau_{j+1} - \tau_j) \left\|\tilde{\theta}_{\tau'_{k(j)}} - \tilde{\theta}_{\tau_j}\right\|^2.
\end{aligned}
$$

Proof. By combining Lemmas 1, 8 and 9 we obtain the following upper bound

$$\begin{aligned}L_{\mathrm{BF}} - L_{\mathrm{Dyn},\boldsymbol{\tau}} \leq & \frac{d}{2}\epsilon_1^2 + \frac{1}{2}\left(\tilde{\theta}_1 - \theta_{\mathrm{init}}\right)^\top \Sigma_{\mathrm{init}}^{-1}\left(\tilde{\theta}_1 - \theta_{\mathrm{init}}\right) - d\log\epsilon_1 - \frac{d}{2}|\boldsymbol{\tau}'| + (|\boldsymbol{\tau}'|-1)d\log\delta \\ & + \sum_{t=2}^{|\boldsymbol{\tau}'|}\left[\frac{1}{2\delta^2}\left(d\left(\epsilon_{t-1}^2 + \epsilon_t^2\right) + \left(\tilde{\theta}_{\tau'_t} - \tilde{\theta}_{\tau'_{t-1}}\right)^\top \Sigma_{\mathrm{init}}^{-1}\left(\tilde{\theta}_{\tau'_t} - \tilde{\theta}_{\tau'_{t-1}}\right)\right) - d\log\epsilon_t\right] \\ & - \log p_0(\boldsymbol{\tau}') + \frac{1}{2}cnR^2\sum_{j=1}^{|\boldsymbol{\tau}|}(\tau_{j+1}-\tau_j)\left(\epsilon_{k(j)}^2\operatorname{Tr}(\Sigma_{\mathrm{init}}) + \left\|\tilde{\theta}_{\tau'_{k(j)}} - \tilde{\theta}_{\tau_j}\right\|^2\right).\end{aligned} \tag{40}$$

We minimize the upper bound with respect to $(\epsilon_j)_{j=1,2,\ldots,|\boldsymbol{\tau}'|}$.
For $j=1$, ϵ_1 only contributes to the above bound through the terms

$$\frac{1}{2}\left(d + \frac{d}{\delta^2} + cnR^2\operatorname{Tr}(\Sigma_{\mathrm{init}})\left(\tau'_2 - \tau'_1\right)\right)\epsilon_1^2 - d\log\epsilon_1. \tag{41}$$

For $j = 2, ..., |\tau'|-1$, ϵ_j only contributes to the bound through the terms

$$\frac{1}{2}\left[\frac{2d}{\delta^2} + cnR^2\operatorname{Tr}(\Sigma_{\mathrm{init}})\left(\tau'_{j+1} - \tau'_j\right)\right]\epsilon_j^2 - d\log\epsilon_j. \tag{42}$$

For $j = |\boldsymbol{\tau}'|$, $\epsilon_{|\boldsymbol{\tau}'|}$ only contributes to the bound through the terms

$$\frac{1}{2}\left[\frac{d}{\delta^2} + cnR^2\operatorname{Tr}(\Sigma_{\mathrm{init}})\left(\tau'_{|\boldsymbol{\tau}'|+1} - \tau'_{|\boldsymbol{\tau}'|}\right)\right]\epsilon_{|\boldsymbol{\tau}'|}^2 - d\log\epsilon_{|\boldsymbol{\tau}'|}. \tag{43}$$

It follows that the upper bound is minimized for

$$\epsilon_1^2 = \frac{d}{d + \frac{d}{\delta^2} + cnR^2\operatorname{Tr}(\Sigma_{\mathrm{init}})\left(\tau'_2 - \tau'_1\right)}, \tag{44}$$

$$\epsilon_j^2 = \frac{d}{\frac{2d}{\delta^2} + cnR^2\operatorname{Tr}(\Sigma_{\mathrm{init}})\left(\tau'_{j+1} - \tau'_j\right)}, \quad \forall j \in \{2,3,\ldots,|\boldsymbol{\tau}'|-1\}, \tag{45}$$

$$\epsilon_{|\boldsymbol{\tau}'|} = \frac{d}{\frac{d}{\delta^2} + cnR^2\operatorname{Tr}(\Sigma_{\mathrm{init}})\left(\tau'_{|\boldsymbol{\tau}'|+1} - \tau'_{|\boldsymbol{\tau}'|}\right)}. \tag{46}$$

Note that the upper bound (40) is a sum of the terms (41), (42) repeated once for each $j \in \{2,3,\ldots,|\boldsymbol{\tau}'|-1\}$, (43), and the following remaining terms

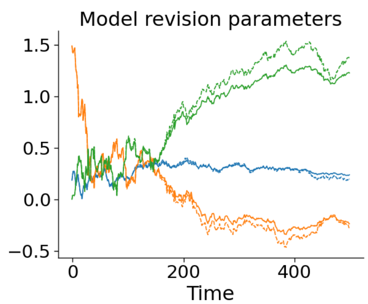
$$\begin{aligned}&\frac{1}{2}\left(\tilde{\theta}_1 - \theta_{\mathrm{init}}\right)^\top \Sigma_{\mathrm{init}}^{-1}\left(\tilde{\theta}_1 - \theta_{\mathrm{init}}\right) - \frac{d}{2}|\boldsymbol{\tau}'| + (|\boldsymbol{\tau}'|-1)d\log\delta - \log p_0(\boldsymbol{\tau}') \\ &+ \sum_{t=2}^{|\boldsymbol{\tau}'|}\frac{1}{2\delta^2}\left(\tilde{\theta}_{\tau'_t} - \tilde{\theta}_{\tau'_{t-1}}\right)^\top \Sigma_{\mathrm{init}}^{-1}\left(\tilde{\theta}_{\tau'_t} - \tilde{\theta}_{\tau'_{t-1}}\right) + \frac{1}{2}cnR^2\sum_{j=1}^{|\boldsymbol{\tau}|}(\tau_{j+1}-\tau_j)\left\|\tilde{\theta}_{\tau'_{k(j)}} - \tilde{\theta}_{\tau_j}\right\|^2.\end{aligned}$$

Plugging in (44), (45) and (46), we get the desired bound. □

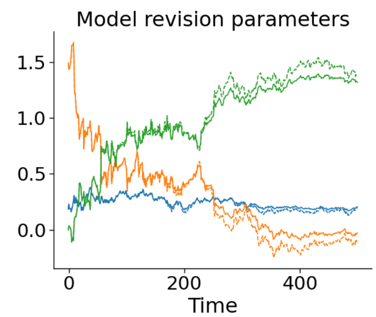
References

K Gordon and A F M Smith. Modeling and monitoring biomedical time series. *J. Am. Stat. Assoc.*, 85(410): 328–337, June 1990. URL `https://doi.org/10.1080/01621459.1990.10476205`.

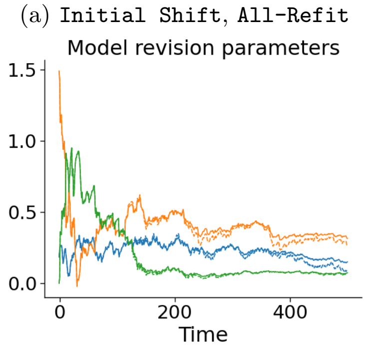
U Orguner and M Demırekler. Analysis of single gaussian approximation of gaussian mixtures in bayesian filtering applied to mixed multiple-model estimation. *Int. J. Control*, 80(6):952–967, June 2007. URL `https://doi.org/10.1080/00207170701261952`.



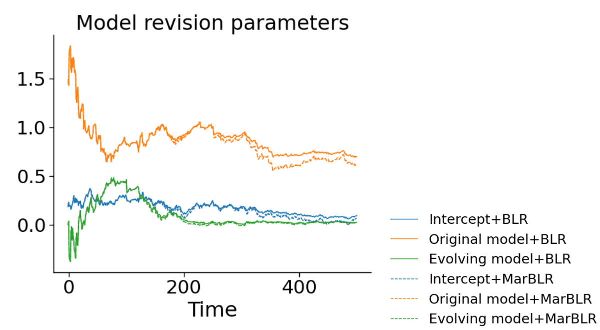
(a) `Initial Shift`, `All-Refit`



(b) `Decay`, `All-Refit`

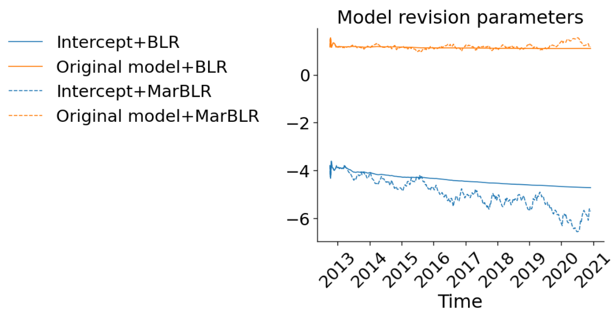


(c) `Initial Shift`, `Subset-Refit`

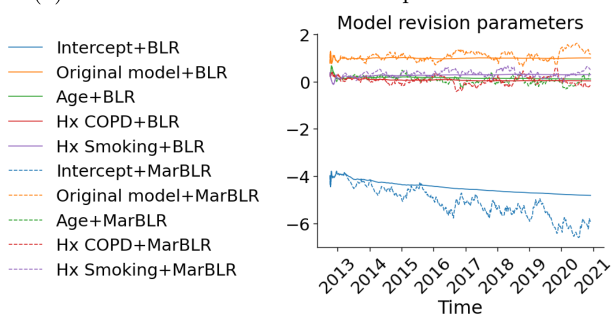


(d) `Decay`, `Subset-Refit`

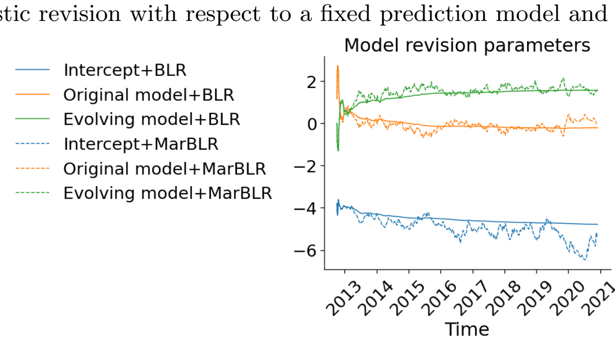
Figure A.1: Evolution of the estimated intercepts and coefficients by BLR and MarBLR when combining the original model with an evolving prediction model (Scenario 3). Data is simulated to be stationary over time after an initial shift (`Initial Shift`) and nonstationary such that the original model decays in performance over time (`Decay`). Underlying prediction model is updated by continually refitting on all previous data (`All-Refit`) or refit on the most recent subset of data (`Subset-Refit`).



(a) Online recalibration of a fixed prediction model



(b) Online logistic revision with respect to a fixed prediction model and patient variables



(c) Online ensembling of the original and continually-refitted prediction models

Figure A.2: Evolution of the estimated intercepts and coefficients for online recalibration and revision of a fixed COPD risk prediction model (a and b, respectively) and online reweighting for fixed and continually-refitted (evolving) COPD risk prediction models using BLR and MarBLR.

Nicholas G Polson, James G Scott, and Jesse Windle. Bayesian inference for logistic models using Pólya–Gamma latent variables. *J. Am. Stat. Assoc.*, 108(504):1339–1349, December 2013. URL `https://doi.org/10.1080/01621459.2013.829001`.

Mike West and Jeff Harrison. *Bayesian Forecasting and Dynamic Models*. Springer, New York, NY, 1997. URL `https://link.springer.com/book/10.1007%2Fb98971`.