A Appendix

A.1 Steps for the updating scheme of Bernoulli responses

Let $\eta = \pi$ and $y|\pi \sim Ber(\pi)$, $0 < \pi < 1$. The conjugate prior is a beta density $q(\pi|\boldsymbol{\tau} = exp\{\tau_1 \ln(\frac{\pi}{1-\pi}) + \tau_0 \ln(1-\pi) - [\ln(\Gamma(\tau_1+1)) + \ln(\Gamma(\tau_0-\tau_1+1)) - \ln(\Gamma(\tau_0+2))]\}$. Consider the linear predictor $\lambda = logit(\pi) = \boldsymbol{F}'\boldsymbol{\theta}$. Following Algorithm **??** :

- Step 1: Given posterior moments for the states, obtain the prior moments of $\lambda \sim N(f, Q)$.
- Step 2.1: Obtain the conjugate prior sufficient statistics vector: $\boldsymbol{H}'_q(\pi) = (\ln(\frac{\pi}{1-\pi}), \ln(1-\pi))$.
- Step 2.2: Solve the system $E_q[\boldsymbol{H}_q(\pi)] = E_p[\boldsymbol{H}_q(\pi)]$, where p is the prior density induced by the normal specification for the linear predictors:

$$\begin{cases} \gamma(\tau_1+1) - \gamma(\tau_0-\tau_1+1) & = f \\ \gamma(\tau_0-\tau_1+1) - \gamma(\tau_0+2) & \simeq \ln(\frac{1}{1+e^f}) - \frac{qe^f}{(1+e^f)^2}. \end{cases}$$

- Step 3: Update the hyperparameters of the conjugate specification: $\tau_1^* = \tau_1 + y, \tau_0^* = \tau_0 + 1$.
- Step 4.1: Let q denote a normal density for the linear predictor, with sufficient statistics vector: $\boldsymbol{H}'_q = (\lambda,\ \lambda^2)$.
- Step 4.2: Solve the system $E_q[\boldsymbol{H}_q(\boldsymbol{\lambda})] = E_p[\boldsymbol{H}_q(\boldsymbol{\lambda})]$, where p is the updated distribution, obtaining: $f^* \simeq \gamma(\tau_1^*+1) - \gamma(\tau_0^*-\tau_1^*+1)$ and $q^* = \gamma'(\tau_1^*+1) + \gamma(\tau_0^*-\tau_1^*+1)$.
- Step 5: The updated moments of the states $\boldsymbol{\theta}$ are trivially obtained, applying normal distribution properties.

The predictive distribution for y_t is a Binomial-Beta$(\tau_0, \tau_1, 1)$, which can be evaluated once step 2.2 is completed.

A second-order Taylor approximation was used to explicitly obtain the system in step 2.2. Also in step 2.2, the system was solved using Newton-Raphson method, with negligible computational cost.

A.2 Analytical development of the proposed method for normal and gamma members of the exponential family

This appendix provides detailed description of the analytical developments involved in the process of information updating for the normal and gamma distributions within the exponential family. These are the sole instances, among the ones presented, in which the application of the suggested methodology is not straightforward. The outlined procedures are executed at each time point t during the information updating process.

Normal distribution with unknown mean and precision: Consider the general formulation and parametrization defined in Section **??** for the normal particular case, with dynamic predictive structure for the mean and precision. Since $\boldsymbol{\lambda} \sim N_2(\boldsymbol{f}, \boldsymbol{Q})$, a system of equations in τ remains to be solved, that is, $E_q[\boldsymbol{H}_q(\mu, \phi)] = E_p[\boldsymbol{H}_q(\mu, \phi)]$, with q denoting a normal-gamma prior distribution and p, the density implied by the joint Gaussianity of the linear predictors (**??**). Using the facts that: i) $E_q[\boldsymbol{H}_q(\mu, \phi)] = \nabla b(\tau_1, \tau_2, \tau_3, \tau_0)$; ii) $E_p[\boldsymbol{H}_q(\mu, \phi)] = E_{p(\phi)}[E_{p(\mu|\phi)}\boldsymbol{H}_q(\mu, \phi)]$ which can be obtained by normal conditioning properties; iii) under p, $\phi \sim LN(f_2, q_2)$ and $\ln \phi \sim N(f_2, q_2)$, results in the system:

$$\begin{cases} \frac{(2\tau_0+1)\tau_2^2}{2\tau_1\tau_2^2 - 8\tau_1^2\tau_3} - \frac{1}{2\tau_1} & = \exp(f_2 + q_2/2)[(f_1 + q_{12})^2 + q_1] \\ \frac{-(2\tau_0+1)\tau_2}{\tau_2^2 - 4\tau_1\tau_3} & = \exp(f_2 + q_2/2)(f_1 + q_{12}) \\ \frac{4\tau_1(\tau_0+1/2)}{\tau_2^2 - 4\tau_1\tau_3} & = \exp(f_2 + q_2/2) \\ \gamma(\tau_0 + 1/2) - \ln\left(\frac{\tau_2^2}{4\tau_1} - \tau_3\right) & = f_2 \end{cases}, \qquad (A.1)$$

with $\gamma(\cdot)$ denoting the digamma function. This system can be analytically solved, as follows: applying the reparametrization in (**??**) to the third equation in (A.1), it follows that $n_0/d_0 = \exp(f_2 + q_2/2)$. Multiplying the third equation by $-\frac{2\tau_2}{4\tau_1} = \mu_0$ we obtain the second equation in (A.1) resulting in $\mu_0 = f_1 + q_{12}$. Doing some algebra in the first equation of (A.1), it follows that $c_0 = [q_1 \exp(f_2 + q_2/2)]^{-1}$. Finally, applying the approximation $\gamma(u) \approx \ln(u) - 1/(2u)$ to the fourth equation of the system, results in $n_0 = 2/q_2$. Applying the reparametrization in (**??**) we obtain the following closed analytical expressions:

$$\begin{cases} \tau_0 &= 1/q_2 - 1/2 \\ \tau_1 &= -[2q_1 \exp(f_2 + q_2/2)]^{-1} \\ \tau_2 &= (f_1 + q_{12})[q_1 \exp(f_2 + q_2/2)]^{-1} \\ \tau_3 &= -(f_1 + q_{12})^2[2q_1 \exp(f_2 + q_2/2)]^{-1} - [q_2 \exp(f_2 + q_2/2)]^{-1}. \end{cases}.$$

After observing $\boldsymbol{y}$, we obtain $\boldsymbol{\tau}^* = (\tau_1^*, \tau_2^*, \tau_3^*, \tau_0^*) = (\tau_1 - 1/2, \tau_2 + y, \tau_3 - y^2/2, \tau_0 + 1/2)$, the updated canonical parameters of a normal-gamma posterior density for (μ, ϕ), and need to evaluate the parameters $(\boldsymbol{f}^*, \boldsymbol{Q}^*)$ of the posterior distribution of the linear predictors that are compatible with them, solving: $E_q[\boldsymbol{H}_q] = E_p[\boldsymbol{H}_q]$, now considering that p is a normal-gamma density for (μ, ϕ) and q is a bivariate normal for $(\mu, \ln\phi)$. $\boldsymbol{H}_q' = (\boldsymbol{\lambda}, \boldsymbol{\lambda}\boldsymbol{\lambda}')$, so the following system is trivially solved:

$$\begin{cases} f_1^* &= -\frac{\tau_2^*}{2\tau_1^*} \\ f_2^* &= \gamma(\tau_0^* + 1/2) - \ln\left(\frac{(\tau_2^*)^2}{4\tau_1^*} - \tau_3^*\right) \\ Q_{11}^* &= \frac{\tau_2^{*2} - 4\tau_1\tau_3}{4\tau_0\tau_1 - 2\tau_1} \\ Q_{12}^* &= 0 \\ Q_{22}^* &= \gamma'(\tau_0^* + 1/2). \end{cases}$$

Notice that, no matter the values of $\tau_0^*, \tau_1^*, \tau_2^*$ and τ_3^*, $Q_{12}^* = 0$, since the parameters that minimize the KL divergence from p to q must satisfy $Cov_p[\mu, \ln(\phi)] = Cov_q[\mu, \ln(\phi)]$ and, as a property of the normal-gamma distribution, we have that (μ, ϕ) are uncorrelated (although not independent), consequently, it can be shown that $(\mu, \ln\phi)$ are also uncorrelated.

Using the reparametrization described in (**??**) we obtain:

$$\begin{cases} f_1^* &= \mu_0^* \\ f_2^* &= \gamma(n_0^*/2) - \ln(d_0^*/2) \\ Q_{11}^* &= \frac{d_0^*/2}{c_0^*(n_0^*/2 - 1)} \\ Q_{12}^* &= 0 \\ Q_{22}^* &= \gamma'(n_0^*/2). \end{cases}$$

An important point to discuss is the equation that defines Q_{11}^*. Specifically, notice that Q_{11}^* is not defined for $n_0^* \leq 2$. Indeed, by the Projection Theorem, the parameters that minimize the KL divergence from p to q must satisfy $Var_p[\mu] = Var_q[\mu]$, but in the conjugated posterior distribution $\mu|\phi \sim N(\mu_0^*, (c_0^*\phi)^{-1})$ and $\phi \sim \mathcal{G}(n_0^*/2, d_0^*/2)$, which implies that $\mu \sim t\left(n_0^*, \mu_0^*, \frac{d_0^*/2}{c_0^* n_0^*/2}\right)$, so that when $n_0^* \leq 2$, $Var_p[\mu] = +\infty$, and the system $E_q[\boldsymbol{H}_q] = E_p[\boldsymbol{H}_q]$ has no valid solutions. Indeed, one can always reduce the KL divergence from p to q by increasing Q_{11}^*. This situation is quite undesirable since it compromises the use of our proposed methodology. Yet, we can restrict the parameter space for the normal posterior distribution to guarantee that, inside that restricted space, we always have a minimum for the divergence from p to q. A natural restriction for the parameter space is to set $Q_{11}^* = \frac{d_0^*/2}{c_0^* n_0^*/2}$, since:

- this restriction guarantees that the scale parameter for the marginal distribution of μ is identical in both the normal and the conjugated distribution.

- for a large value of n_0^* (which we expect to have for a reasonable sample size), $\frac{d_0^*/2}{c_0^* n_0^*/2} \approx \frac{d_0^*/2}{c_0^*(n_0^*/2-1)}$, i.e., for large values of n_0^* this restriction has no significant effect, in the sense that the optimum in the restricted space will be very close to the global optimum.

- $Q_{11}^* = \frac{d_0^*/2}{c_0^* n_0^*/2}$ becomes numerically well behaved for any possible value of n_0^*, since, after updating our knowledge of μ, it is guaranteed that $n_0^* \geq 1$, avoiding a division by values close to zero.

Gamma distribution with unknown shape α: Let $\boldsymbol{\eta} = (\alpha, \mu)$ and $(y|\boldsymbol{\eta}) \sim \mathcal{G}(\alpha, \alpha/\mu)$, $\alpha, \mu > 0$. Unlike other particular cases discussed in this work, the normalising constant of the joint prior for the pair (α, μ) is unknown and the conjugate prior is $\pi(\alpha, \mu) \propto \exp\left\{\tau_0\left(\alpha \ln\left(\frac{\alpha}{\mu}\right) - \ln(\Gamma(\alpha)\right) + \tau_1\alpha - \tau_2\frac{\alpha}{\mu}\right\}$. Consider a vector of linear predictors $\boldsymbol{\lambda}' = (\lambda_1, \lambda_2) = (\ln(\mu), \ln(\phi)) = (F_1'\theta_1, F_2'\theta_2)$. Following Algorithm **??**:

- Step 1: Given posterior moments for the states, obtain the prior moments of $\boldsymbol{\lambda} \sim N(\boldsymbol{f}, \boldsymbol{Q})$.

- Step 2.1: Obtain the conjugate prior sufficient statistics vector:
  $\boldsymbol{H}_q = \left(\alpha\ln(\mu) - \alpha\ln(\alpha) + \ln(\Gamma(\alpha)), \alpha, \frac{\alpha}{\mu}\right)'$.

- Step 2.2: Solve the system $E_q[\boldsymbol{H}_q(\boldsymbol{\eta})] = E_p[\boldsymbol{H}_q(\boldsymbol{\eta})]$, where p is the prior density induced by the normal specification for the linear predictors. It is easily seen that $\mu|\alpha \sim \mathcal{IG}(\tau_0\alpha - 1, \alpha\tau_2)$ where $\mathcal{IG}$ denotes the inverse gamma density, if $\alpha > 1/\tau_0$. Using the fact that $E_q[\boldsymbol{H}_q] =$

$E_q[E_q[\boldsymbol{H}_q|\alpha]]$ and that $E_q[\ln(\mu)|\alpha] = -\gamma(\tau_0\alpha - 1) + \ln(\alpha\tau_2)$, it follows that:

$$\begin{aligned} E_q\left[\frac{\alpha}{\mu}\right] &= E_q\left[\alpha E_q\left[\frac{1}{\mu}|\alpha\right]\right] = E_q\left[\alpha\frac{\tau_0\alpha+1}{\alpha\tau_2}\right] = \frac{\tau_0 E_q\left[\alpha\right]-1}{\tau_2}; \\ E_q[\alpha\ln(\mu)] &= E_q[\alpha E_q[\ln(\mu)|\alpha]] = E_q\left[\alpha(-\gamma(\tau_0\alpha+1)+\ln(\alpha\tau_2))\right]. \end{aligned}$$

Assuming $E_p\left[H_q\right]$ known, and using the fact that $E_q[\alpha] = E_p[\alpha]$ (from the system to be solved), it follows that $\tau_2 = \frac{\tau_0 E_q[\alpha]-1}{E_q\left[\frac{\alpha}{\mu}\right]}$. One can trivially find that $E_q[\alpha\ln(\mu) - \alpha\ln(\alpha) + \ln(\Gamma(\alpha)] = E_q\left[\ln(\Gamma(\alpha) - \alpha\gamma(\tau_0\alpha+1)\right] + E_q\left[\alpha\right]\ln(\tau_2)$. Thus, $E_q[\boldsymbol{H}_q]$ can be expressed in terms of expected values which solely depend on the marginal prior for α: $\pi(\alpha) = \int_0^{+\infty}\pi(\alpha,\mu)d\mu \propto \frac{\alpha^{\tau_0\alpha}}{\alpha^{-\tau_0}\Gamma(\alpha+1)^{\tau_0}}\frac{\Gamma(\tau_0\alpha+1)}{\alpha^{\tau_0\alpha+1}\tau_2^{\tau_0\alpha+1}}\exp\{\tau_1\alpha\}$. Applying Stirling's formula $\Gamma(x+1) \approx \sqrt{2\pi x}\left(\frac{x}{e}\right)^x$ to this last expression results in: $\pi(\alpha) \propto \alpha^{\frac{\tau_0-1}{2}-1}\exp\{-\left[\tau_0\ln\left(\frac{\tau_2}{\tau_0}\right)-\tau_1\right]\alpha\}$, that is, α approximately follows a $\mathcal{G}\left(\frac{\tau_0+1}{2}, \tau_0\ln\left(\frac{\tau_2}{\tau_0}\right)-\tau_1\right)$ marginal conjugate prior, so that:

$$E_q[\alpha] = \frac{\tau_0+1}{2\left(\tau_0\ln\left(\frac{\tau_2}{\tau_0}\right)-\tau_1\right)}.$$

Using a second order approximation for the digamma function, $\gamma(x) \approx \ln(x) - \frac{1}{2x} - \frac{1}{12x^2}$, and the fact that $\gamma(x+1) = \gamma(x) + \frac{1}{x}$ results in:

$$\begin{aligned} -\alpha\gamma(\tau_0\alpha+1) &= -\alpha\left(\gamma(\tau_0\alpha) + \frac{1}{\tau_0\alpha}\right) \\ &\approx -\alpha\left(\ln(\tau_0\alpha) - \frac{1}{2\tau_0\alpha} - \frac{1}{12\tau_0^2\alpha^2} + \frac{1}{\tau_0\alpha}\right) \\ &\approx -\alpha\ln(\alpha) - \alpha\ln(\tau_0) + \frac{1}{12\tau_0^2\alpha} - \frac{1}{2\tau_0}. \end{aligned}$$

Once again using a second order approximation for the digamma function and the fundamental theorem of Calculus:

$$\ln(\Gamma(x)) = \int_1^x \gamma(t)dt \approx \int_1^x \ln(t) - \frac{1}{2t} - \frac{1}{12t^2}dt = x\ln(x) - x - \frac{1}{2}\ln(x) + \frac{1}{12x} + \frac{11}{12}.$$

The stated facts allow us to calculate the expected value of the remaining sufficient statistic, resulting in the following system to be solved:

$$\begin{cases} \mathbb{E}_p\left[(\alpha\ln(\mu) - \alpha\ln(\alpha) + \ln(\Gamma(\alpha))\right] & = \frac{\tau_0+1}{2(\tau_0\ln(\frac{\tau_2}{\tau_0})-\tau_1)}(\ln(\tau_2/\tau_0)-1) - \frac{1}{2}\psi\left(\frac{\tau_0+1}{2}\right) \\ & \quad +\frac{1}{2}\ln\left(\tau_0\ln\left(\frac{\tau_2}{\tau_0}\right)-\tau_1\right) + \frac{\tau_0+1}{\tau_0-1}\frac{\ln\left(\frac{\tau_2}{\tau_0}\right)-\frac{\tau_1}{\tau_0}}{6} \\ & \quad +\frac{11}{12} - \frac{1}{2\tau_0} \\ \exp\{f_1 + Q_{11}/2\} & = \frac{\tau_0+1}{2(\tau_0\ln(\frac{\tau_2}{\tau_0})-\tau_1)} \\ \exp\{f_1 - f_2 + Q_{11}/2 + Q_{12} + Q_{22}/2\} & = \frac{\tau_0(\tau_0+1)-1}{\tau_2 2(\tau_0\ln(\frac{\tau_2}{\tau_0})-\tau_1)} \end{cases}$$

with $\mathbb{E}_p\left[(\alpha\ln(\mu) - \alpha\ln(\alpha) + \ln(\Gamma(\alpha))\right]$ evaluated by numerical integration, with negligible cost.

- Step 3: Update the hyperparameters of the conjugate specification: $\boldsymbol{\tau}^* = (\tau_0^*, \tau_1^*, \tau_2^*) = (\tau_0+1, \tau_1+\ln(y), \tau_2+y)$.

- Step 4.1: Let q denote a multivariate normal density for the vector of linear predictors, with sufficient statistics vector: $\boldsymbol{H}_q' = (\boldsymbol{\lambda},\ \boldsymbol{\lambda}\boldsymbol{\lambda}')$.

- Step 4.2: Solve the system $E_q[\boldsymbol{H}_q(\boldsymbol{\lambda})] = E_p[\boldsymbol{H}_q(\boldsymbol{\lambda})]$, where p is the updated distribution, obtaining:: $f_1^* = E_p[\ln(\mu)]$, $f_2^* = E_p[\ln(\alpha)]$; $Q_{11}^* = V_p[\ln(\mu)]$, $Q_{12}^* = Cov_p[\ln(\mu), \ln(\alpha)]$, $Q_{22}^* = E_p[\ln(\alpha)]$. Using the fact that α approximately follows a $\mathcal{G}\left(\frac{\tau_0^*+1}{2}, \tau_0^*\ln\left(\frac{\tau_2^*}{\tau_0^*}\right) - \tau_1^*\right)$ conjugate marginal posterior, it follows that:

$$\begin{aligned} E_p[\ln(\alpha)] &\approx \gamma\left(\frac{\tau_0^*+1}{2}\right) - \ln\left(\tau_0^*\ln\left(\frac{\tau_2^*}{\tau_0^*}\right) - \tau_1^*\right) \\ V_p[\ln(\alpha)] &\approx \gamma'\left(\frac{\tau_0^*+1}{2}\right). \end{aligned}$$

Since, *a posteriori*, $\mu|\alpha \sim \mathcal{IG}(\tau_0^*\alpha + 1, \alpha\tau_2^*)$, it follows that:

$$\begin{aligned} E_p[\ln(\mu)] &= E_p[E_p[\ln(\mu)|\alpha]] = -E_p\left[\gamma(\tau_0^*\alpha+1) - \ln(\alpha\tau_2^*)\right]; \\ V_p[\ln(\mu)] &= V_p[\ln(1/\mu)] = E_p[V_p[\ln(1/\mu)|\alpha]] + V_p[E_p[\ln(1/\mu)|\alpha]] \\ &= E_p[\gamma(\tau_0^*\alpha+1)] + V_p[\gamma(\tau_0^*\alpha+1) - \ln(\alpha\tau_2^*)]; \\ E_p[\ln(\alpha)\ln(\mu)] &= -E_p[\ln(\alpha)\ln(1/\mu)] = -E_p[\ln(\alpha)E_p[\ln(1/\mu)|\alpha]] \\ &= -E_p[\ln(\alpha)(\gamma(\tau_0^*\alpha+1) - \ln(\alpha\tau_2^*))]. \end{aligned}$$

All the integrals involved in the expected values above are evaluated by numerical integration, with negligible cost.

- Step 5: The updated moments of the states $\boldsymbol{\theta}$ are trivially obtained, applying normal distribution properties.

Since the normalizing constant of the conjugate prior is unknown, we trivially obtain a sample from the predictive distribution, first sampling from the posterior $\boldsymbol{\lambda}$ and then sampling from the observational distribution of $y^*|\boldsymbol{\lambda}$. Since we assume that $\boldsymbol{\lambda}$ follows a multivariate normal distribution, the computational cost of this process is negligible.

A.3 Multinomial model: obtaining the probabilities of hospital admission for each age group.

As seen in Subsection **??**, for a multinomial response on k categories where the $k-$th category is used as a reference group, the following predictive structure is specified:

$$\lambda_{jt} = \ln\left(\frac{\eta_{jt}}{\eta_{k,,t}}\right) = \boldsymbol{F}'_{jt}\boldsymbol{\theta_t}, \quad j = 1, \ldots, k-1.$$

Let E_{jt} denote the exposure of group j; $E_{1:k,t} = \sum_{j=1}^{k} E_{jt}$ denote the total exposure of all age groups and consider the following events:

G_{jt} : allocation of an individual to age group j, $j = 1, \ldots, k-1$; at time t, t=1, 2, …;

H_t: hospital admission of an individual, at time t , t=1, 2, … .

Note that the probability of allocation to age group j and time t, conditionally on having been hospitalized is:

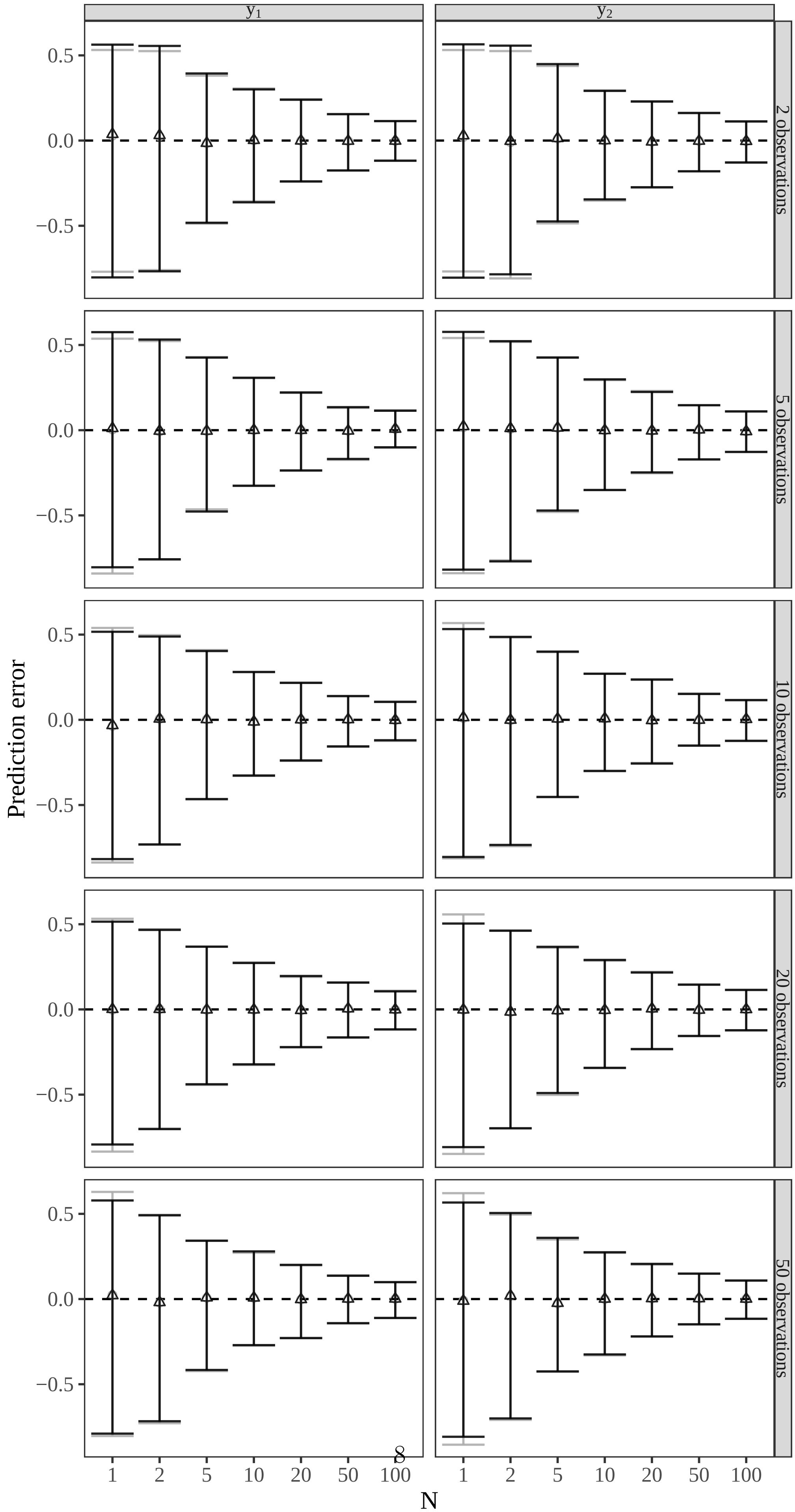
$$\eta_{jt} = \mathbb{P}(G_{jt}|H_t) = \frac{\mathbb{P}(H_t|G_{jt})\mathbb{P}(G_{jt})}{\mathbb{P}(H_t)} \Rightarrow \frac{\mathbb{P}(H_t|G_{jt})}{\mathbb{P}(H_t|G_{kt})} = \frac{\mathbb{P}(G_{jt}|H_t)}{\mathbb{P}(G_{kt}|H_t)}\frac{E_{kt}}{E_{jt}}.$$

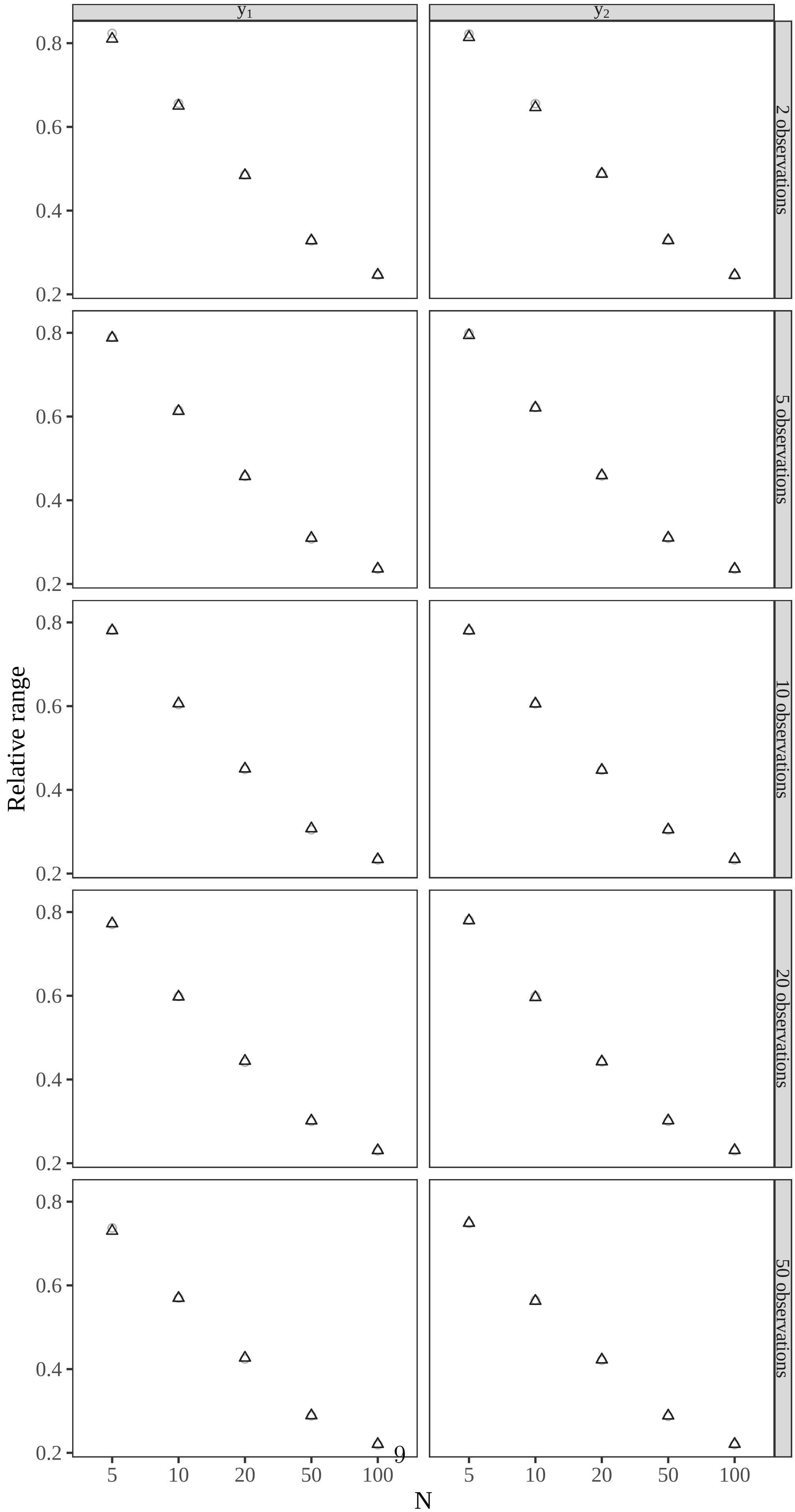
Then it follows that:

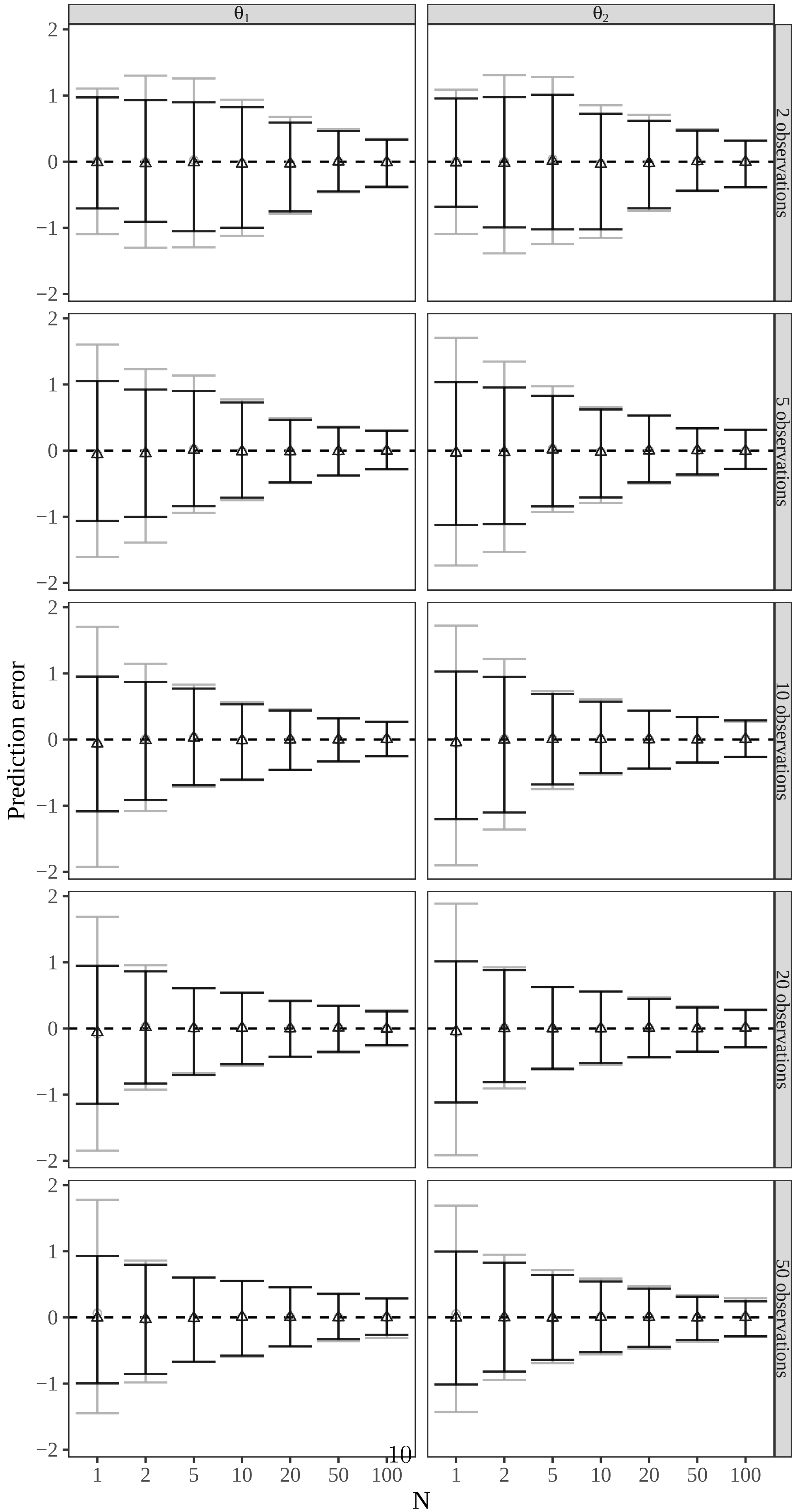
$$\begin{aligned}\ln\left\{\frac{\mathbb{P}(H_t|G_{jt})}{\mathbb{P}(H_t|G_{kt})}\right\} =& \ln\left\{\frac{\mathbb{P}(G_{jt}|H_t)}{\mathbb{P}(G_{kt}|H_t)}\right\} - \ln\left\{\frac{E_{jt}}{E_{kt}}\right\} \\ =& \boldsymbol{F}'_{jt}\boldsymbol{\theta}_t - \ln\left\{\frac{E_{jt}}{E_{kt}}\right\}, \quad j = 1, ..., k-1.\end{aligned}$$

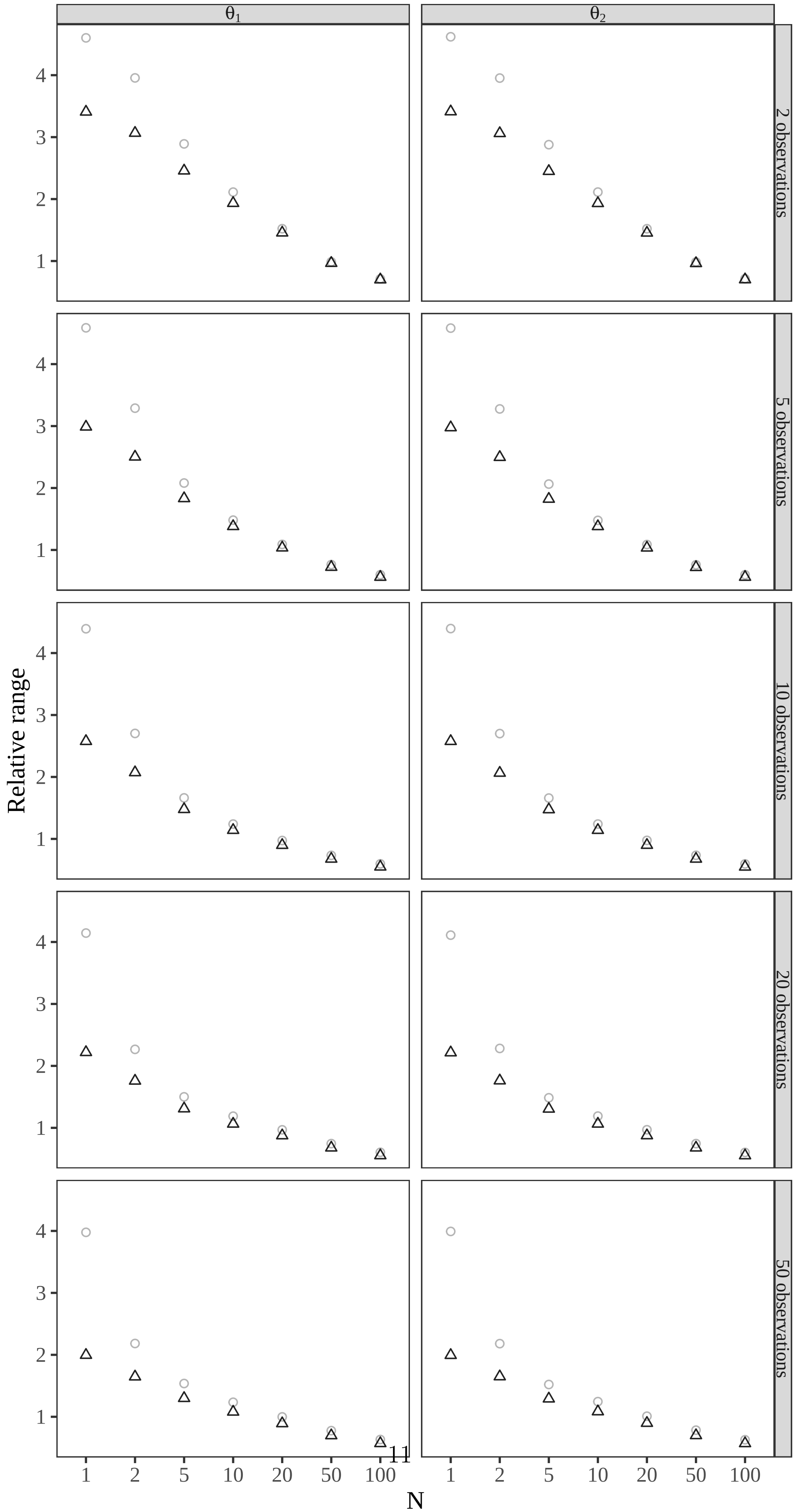
Thus, estimates of $\mathbb{P}(H_t|G_{jt})$ are naturally obtained if $\ln\left\{\frac{E_{jt}}{E_{kt}}\right\}$ is introduced as an offset term in the dynamic predictor of category j, $j = 1, ..., k-1$.

A.4 Detailed figures for the simulated example


y1
y2
2 observations
5 observations
10 observations
20 observations
50 observations
Prediction error
0.5
0.0
−0.5
1
2
5
10
20
50
100
N


y1
y2
2 observations
5 observations
10 observations
20 observations
50 observations
Relative range
0.2
0.4
0.6
0.8
5
10
20
50
100
N


θ1
θ2
2 observations
5 observations
10 observations
20 observations
50 observations
Prediction error
N
1
2
5
10
20
50
100
2
1
0
−1
−2


θ1
θ2
2 observations
5 observations
10 observations
20 observations
50 observations
Relative range
N
1
2
5
10
20
50
100
1
2
3
4