

## Model Setup

Consider a standard OLS regression applied to the dataset  $d_k = \{y_{i,k}, x_{i1,k}, \dots, x_{ip,k}\}_{i=1}^n$  in the test of anomaly  $k$

$$y_{i,k} = \beta_{0,k} + \beta_{1,k}x_{i1,k} + \dots + \beta_{p,k}x_{ip,k} + \varepsilon_{i,k} \quad (1)$$

with independent residual errors  $\varepsilon_{i,k} \sim \mathcal{N}(0, \sigma^2)$ <sup>1</sup>. In this setup, let  $\mathcal{M}_{\text{null},k}$  denote the null model that restricts  $\beta_{0,k} \leq 0$

$$\mathcal{M}_{\text{alt},k} := \{y_{i,k} = \beta_{0,k} + \beta_{1,k}x_{i1,k} + \dots + \beta_{p,k}x_{ip,k} + \varepsilon_{i,k}, \quad \beta_{0,k} \leq 0\}$$

and  $\mathcal{M}_{\text{alt},k}$  denote the alternative model that restricts  $\beta_{0,k} > 0$

$$\mathcal{M}_{\text{alt},k} := \{y_{i,k} = \beta_{0,k} + \beta_{1,k}x_{i1,k} + \dots + \beta_{p,k}x_{ip,k} + \varepsilon_{i,k}, \quad \beta_{0,k} > 0.\}$$

Our goal is to compute the posterior probability of the null model,  $\Pr(\mathcal{M}_{\text{null},k} | d_k)$ , given the data  $d_k$ .

We place the non-informative (Jeffreys) prior on the other coefficients  $\boldsymbol{\beta}_{1:p,k} = (\beta_{1,k}, \dots, \beta_{p,k})^\top$  and the error variance  $\sigma^2$

$$\pi(\boldsymbol{\beta}_{1:p,k}, \sigma^2) \propto \frac{1}{\sigma^2}.$$

For  $\beta_{0,k}$ , we use a mixed one-sided prior derived from the base normal distribution  $\mathcal{N}(0, \tau^2)$ , where  $\tau$  is the prior scale parameter, that encodes a belief about its sign

$$\pi(\beta_{0,k}) = \pi_{\text{null}} 2\tau^{-1} \phi(\beta_{0,k} / \tau) \mathbb{I}(\beta_{0,k} \leq 0) + (1 - \pi_{\text{null}}) 2\tau^{-1} \phi(\beta_{0,k} / \tau) \mathbb{I}(\beta_{0,k} > 0),$$

where  $\phi(\cdot)$  is the standard normal distribution <sup>2</sup> and  $\mathbb{I}(\cdot)$  is an indicator function equal to one if the condition is satisfied and zero otherwise. The hyperparameter  $0 \leq \pi_{\text{null}} \leq 1$  is the prior probability of  $\mathcal{M}_{\text{null},k}$ , and  $\tau > 0$  is a scale parameter for  $\beta_{0,k}$ . This prior is essentially a normal distribution  $\mathcal{N}(0, \tau^2)$  for  $\beta_{0,k}$  truncated to negative or positive values, with a mixing weight  $\pi_{\text{null}}$  favoring the negative side and  $1 - \pi_{\text{null}}$  the positive side.

---

<sup>1</sup>The assumption of independent model errors is not essential for our framework as models with more sophisticated assumptions on data dependence fit into this framework straightforwardly.

<sup>2</sup>Thus,  $2\tau^{-1}\phi(\beta_{0,k} / \tau)$  is a half-normal distribution on either side of zero.

## Posterior Inference Approach

The full likelihood function for the model (1) conditioned on the data  $d_k$  is

$$L(\beta_{0,k}, \beta_{1:p,k}, \sigma^2 | d_k) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}[\beta_{0,k}, \beta_{1:p,k}])^\top (\mathbf{y} - \mathbf{X}[\beta_{0,k}, \beta_{1:p,k}])\right] \quad (2)$$

where  $\mathbf{y} = \{y_{1,k}, \dots, y_{n,k}\}^\top$ ,  $\mathbf{X} = [\mathbf{1}, \mathbf{Z}]$  with  $\mathbf{1}$  being the column of ones (to accommodate  $\beta_{0,k}$ ) and  $\mathbf{Z}$  containing the  $p$  covariates. Integrate the likelihood function (2) with respect to  $\beta_{1:p,k}$  and  $\sigma^2$  under the Jeffreys prior to obtain the likelihood function for  $\beta_{0,k}$

$$L(\beta_{0,k} | d_k) = c[1 + t^2(\beta_{0,k}) / \nu]^{-\frac{\nu+1}{2}}$$

where

$$t(\beta_{0,k}) = \frac{\beta_{0,k} - \hat{\beta}_{0,k}}{\text{SE}(\hat{\beta}_{0,k})}, \quad \nu = n - p - 1, \quad c = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\text{SE}(\hat{\beta}_{0,k})}(2\pi)^{-\nu/2}|\mathbf{Z}^\top \mathbf{Z}|^{-1/2}.$$

Here,  $t(\beta_{0,k})$  is the usual  $t$ -statistic and  $\text{SE}(\hat{\beta}_{0,k})$  is the standard error of the OLS estimate  $\hat{\beta}_{0,k}$ . Consequently,  $L(\beta_{0,k} | d_k)$  is (up to the constant  $c$ ) the Student  $t$ -distribution for  $\beta_{0,k}$  with  $\nu$  degrees of freedom, centered at  $\hat{\beta}_{0,k}$  and having scale  $\text{SE}(\hat{\beta}_{0,k})$ .

Then, we compute the marginal likelihood under each sub-model,  $L_{\text{null}}(d_k)$  and  $L_{\text{alt}}(d_k)$ , by integrating  $L(\beta_{0,k} | d_k)$  over the half-normal prior on the restricted domain of  $\beta_{0,k}$  for each model, respectively,

$$L_{\text{null}}(d_k) = \int_{-\infty}^0 L(\beta_{0,k} | d_k) 2\tau^{-1} \phi(\beta_{0,k} / \tau) d\beta_{0,k}$$

and

$$L_{\text{alt}}(d_k) = \int_0^{\infty} L(\beta_{0,k} | d_k) 2\tau^{-1} \phi(\beta_{0,k} / \tau) d\beta_{0,k}.$$

The posterior probability of  $\mathcal{M}_{\text{null},k}$  is

$$\Pr(\mathcal{M}_{\text{null},k} | d_k) = \frac{\pi_{\text{null}} L_{\text{null}}(d_k)}{\pi_{\text{null}} L_{\text{null}}(d_k) + (1 - \pi_{\text{null}}) L_{\text{alt}}(d_k)}.$$