

**EXTRA 3 — post\_mlp:  $\text{logit}(\text{'Smith'}) - \text{logit}(\text{'Jones'})$**

