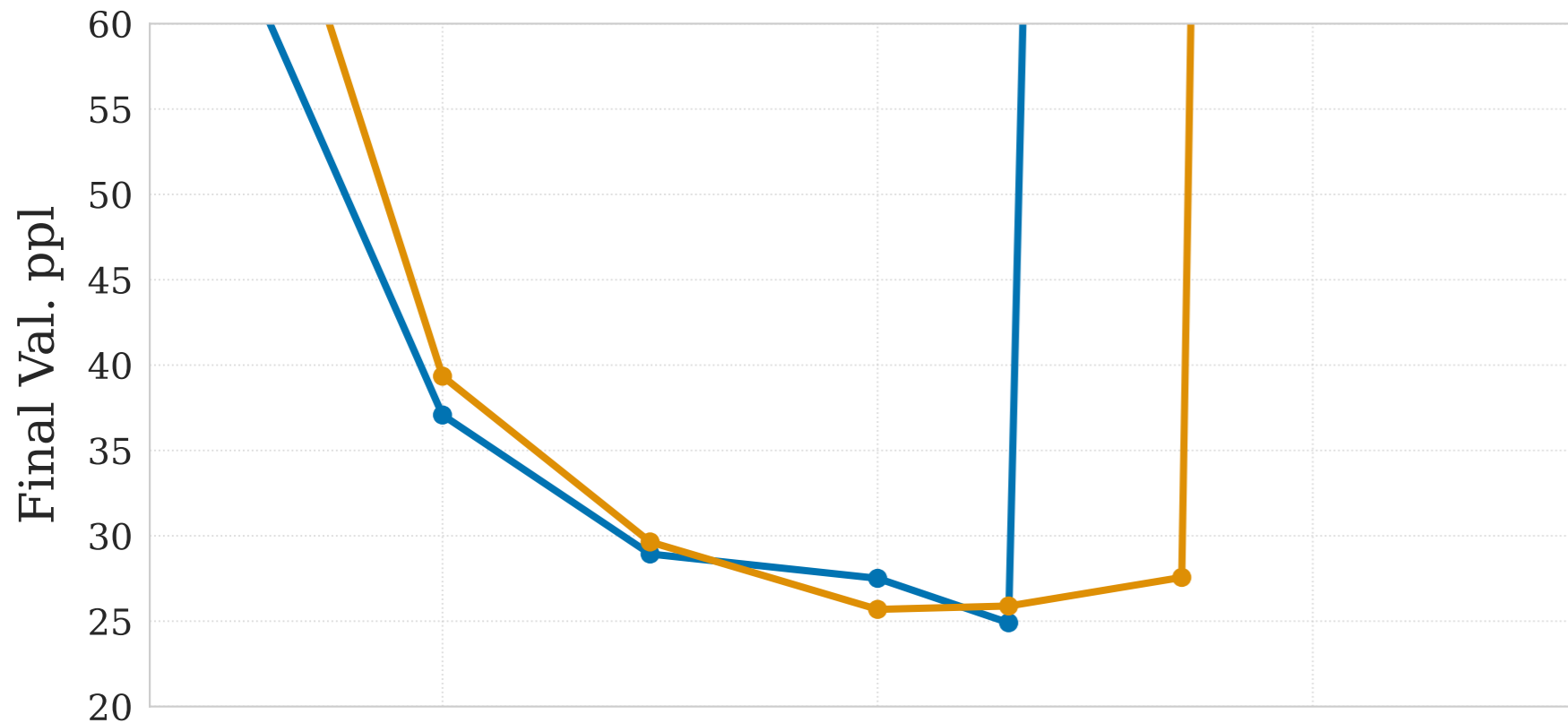
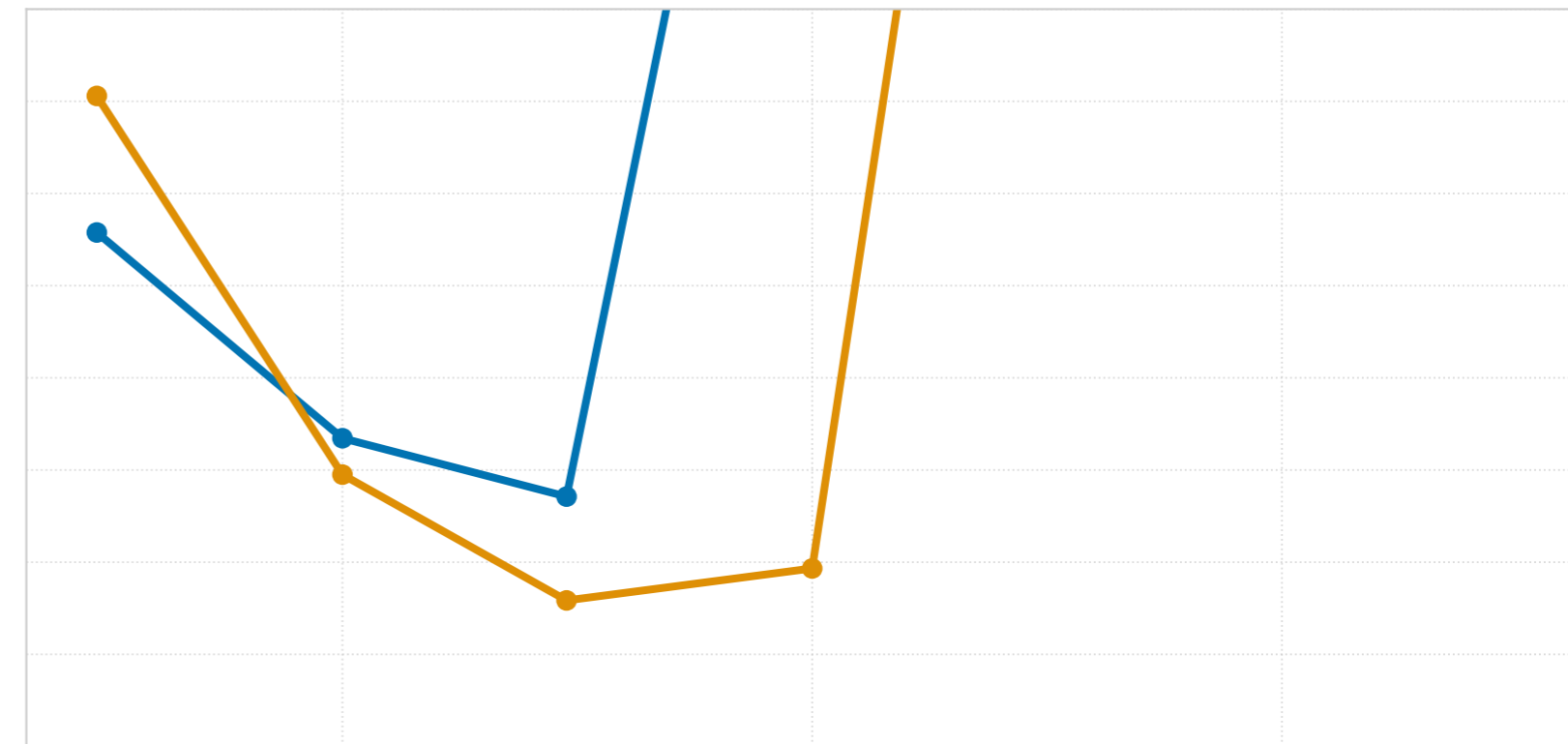


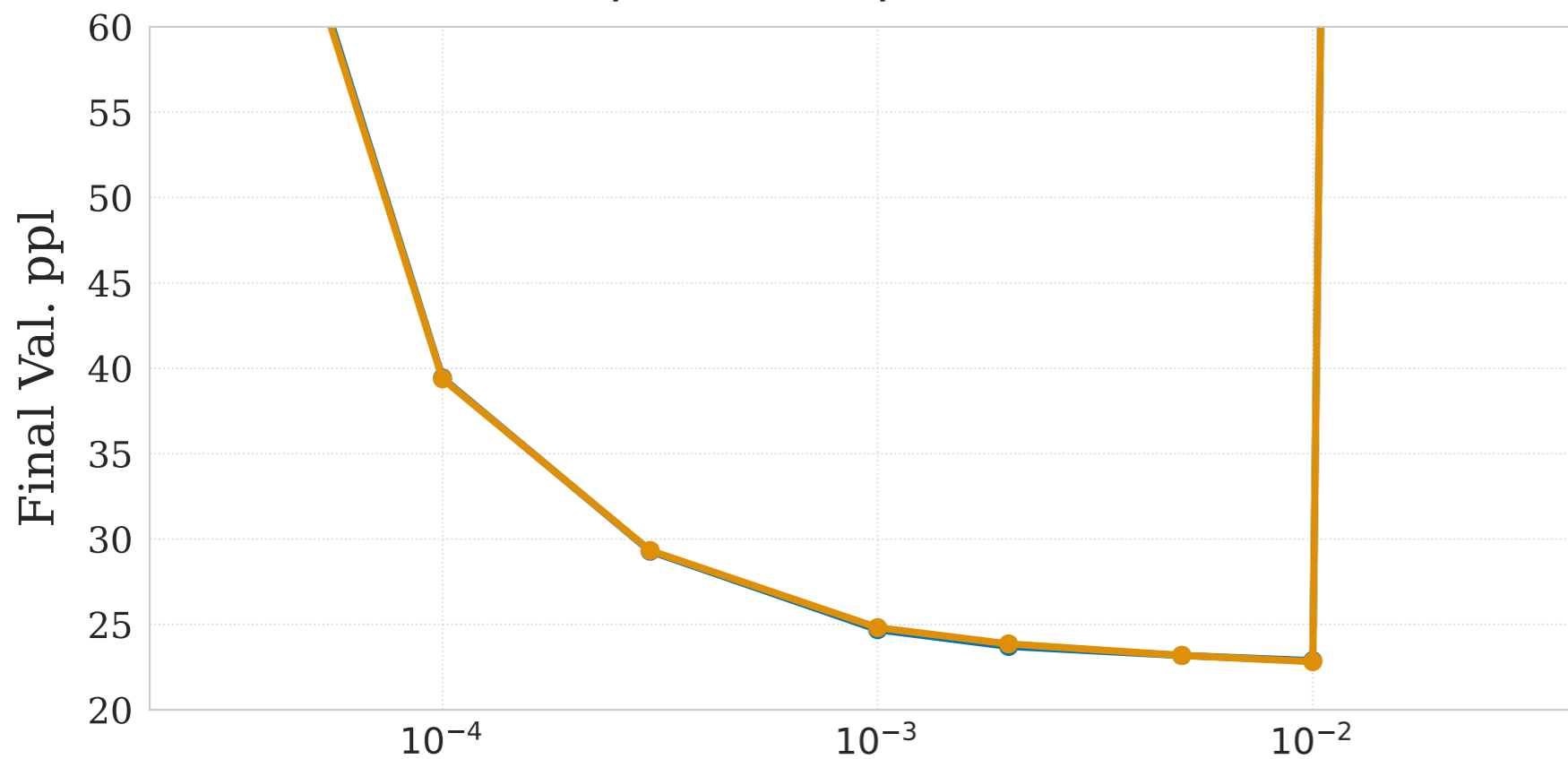
Warmup Cosine  
 $\beta_1 = 0.9, \beta_2 = 0.999$



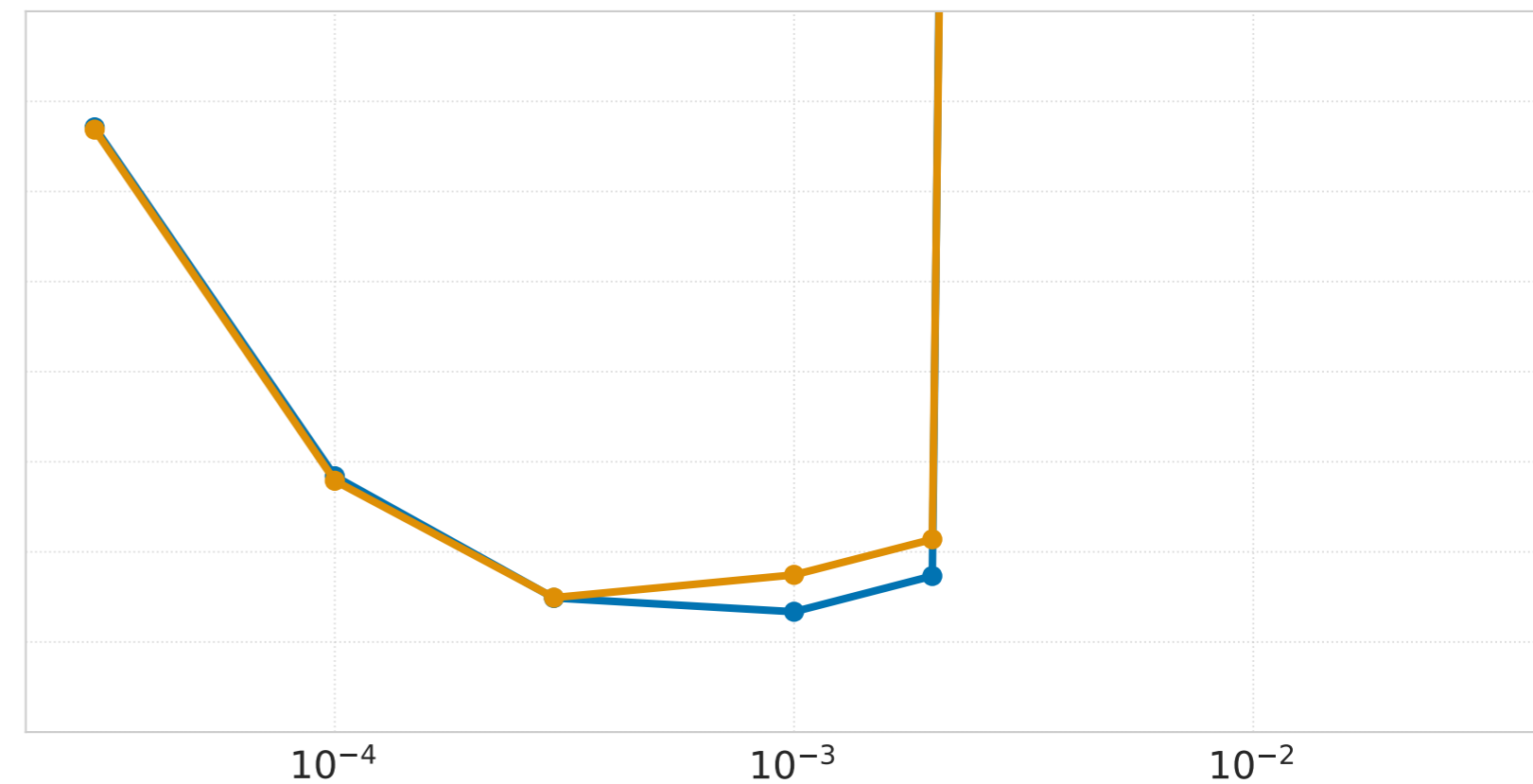
Constant Learning Rate  
 $\beta_1 = 0.9, \beta_2 = 0.999$



$\beta_1 = 0.95, \beta_2 = 0.95$



$\beta_1 = 0.95, \beta_2 = 0.95$



Learning Rate (log scale)

Learning Rate (log scale)

