pre-training in-context prediction X_j X_1 intervention noise abduction