the behavior of γ on the decision boundary fully control the rate of regret decay whenever a cross-validated estimator is used. We also note that, as far as we can tell, there does not appear to be any cost to using an efficient value function estimator when the model is nonparametric. It is not clear if using the efficient influence function is always preferred in more restrictive, semiparametric models: there may need to be a careful tradeoff between the efficiency of the influence function and the corresponding Rem_n.

3.3 Relation to Results of Athey and Wager [1]

In the recent technical report [1], Athey and Wager showed that policy learning regret rates on the order of $O_P(n^{-1/2})$ are attainable by ERMs. High-probability regret upper bounds were derived, with leading constants that scale with the standard error of a semiparametric efficient estimator for policy evaluation. The authors argue that this leading constant demonstrates the importance of using semiparametric efficient value estimators to define the empirical risk used by their estimator. As we discussed in Section 3.2, we fully agree with the importance of using efficient estimators to estimate the empirical risk. Nonetheless, the present work shows that the regret of ERMs decays faster than the standard error of an efficient estimator of the value and so, the regret of ERMs that use this empirical risk will not necessarily scale with the standard error of this estimator.

Like in [1], our results are given under a fixed data generating distribution P. We implicitly leverage the behavior of γ near the decision boundary (zero) under this distribution P. Crucially, there is a problem-dependent constant that can be made arbitrarily large if one chooses a P for which $\gamma(X)$ concentrates a large amount of mass near the decision boundary. Hence, the minimax rate is not faster than $n^{-1/2}$ unless one constrains the class of distributions to which P can belong via a margin condition (see [23, 11]). The implication of this observation is encouraging: it would not be surprising to find that, without a margin condition, the minimax regret does in fact decay at the rate of the standard error of an efficient estimator of an optimal policy. The leading constant may also critically depend on