Our goal is to measure the effectiveness of display advertising. On the other hand, correcting observational data have been proposed to evaluate campaigns in [1].

For reliable cookies, running experiments (A/B testing) has been proposed to evaluate campaigns in [1].

Online display advertising is an area of rapid growth and consequently of great interest as a marketing channel. Achieving statistical significance for campaign lifts is non-trivial for A/B testing as described in [2].

A/B testing is expensive. Can this experimentation be less costly? Achieving statistical significance for campaign lifts is not trivial for commercial campaigns. Can we aggregate prior information to improve these estimations?

We test our method with the PROMO dataset to compare with a ground truth [4]. We use products with less than 6 relevant campaigns for 365 days. We detect 84.8% of effective campaigns correctly, and 73.2% of the days a campaign is effective per product.

Discussion and Current Work

We observed several campaigns with non-significant average effect on actions, which is consistent with A/B testing results. When experimentation is possible and we can track users, can we combine A/B testing with time series?

A/B testing as described in [3] is expensive. Can this experimentation be less costly? Achieving statistical significance for campaign lifts is not trivial for commercial campaigns. Can we aggregate prior information to improve these estimations?

We test two base models: 1. Random Walk (simplest dynamic model) 2. Weekly Seasonal model [4].

We test the model with the number of actions and impressions, (Mu) and using the log transformation to relax the linear relationship between actions and impressions (Mupt).

We analyze 2,885 campaigns associated with 1,251 products during 6 months.

Results

We test two base models: 1. Random Walk (simplest dynamic model) 2. Weekly Seasonal model [4].

We test the model with the number of actions and impressions, (Mu) and using the log transformation to relax the linear relationship between actions and impressions (Mupt).

We analyze 2,885 campaigns associated with 1,251 products during 6 months.

Discussion and Current Work

We observed several campaigns with non-significant average effect on actions, which is consistent with A/B testing results. When experimentation is possible and we can track users, can we combine A/B testing with time series?

A/B testing as described in [3] is expensive. Can this experimentation be less costly? Achieving statistical significance for campaign lifts is not trivial for commercial campaigns. Can we aggregate prior information to improve these estimations?

We test our method with the PROMO dataset to compare with a ground truth [4]. We use products with less than 6 relevant campaigns for 365 days. We detect 84.8% of effective campaigns correctly, and 73.2% of the days a campaign is effective per product.

Discussion and Current Work

We observed several campaigns with non-significant average effect on actions, which is consistent with A/B testing results. When experimentation is possible and we can track users, can we combine A/B testing with time series?

A/B testing as described in [3] is expensive. Can this experimentation be less costly? Achieving statistical significance for campaign lifts is not trivial for commercial campaigns. Can we aggregate prior information to improve these estimations?

We test our method with the PROMO dataset to compare with a ground truth [4]. We use products with less than 6 relevant campaigns for 365 days. We detect 84.8% of effective campaigns correctly, and 73.2% of the days a campaign is effective per product.

Discussion and Current Work

We observed several campaigns with non-significant average effect on actions, which is consistent with A/B testing results. When experimentation is possible and we can track users, can we combine A/B testing with time series?

A/B testing as described in [3] is expensive. Can this experimentation be less costly? Achieving statistical significance for campaign lifts is not trivial for commercial campaigns. Can we aggregate prior information to improve these estimations?

We test our method with the PROMO dataset to compare with a ground truth [4]. We use products with less than 6 relevant campaigns for 365 days. We detect 84.8% of effective campaigns correctly, and 73.2% of the days a campaign is effective per product.