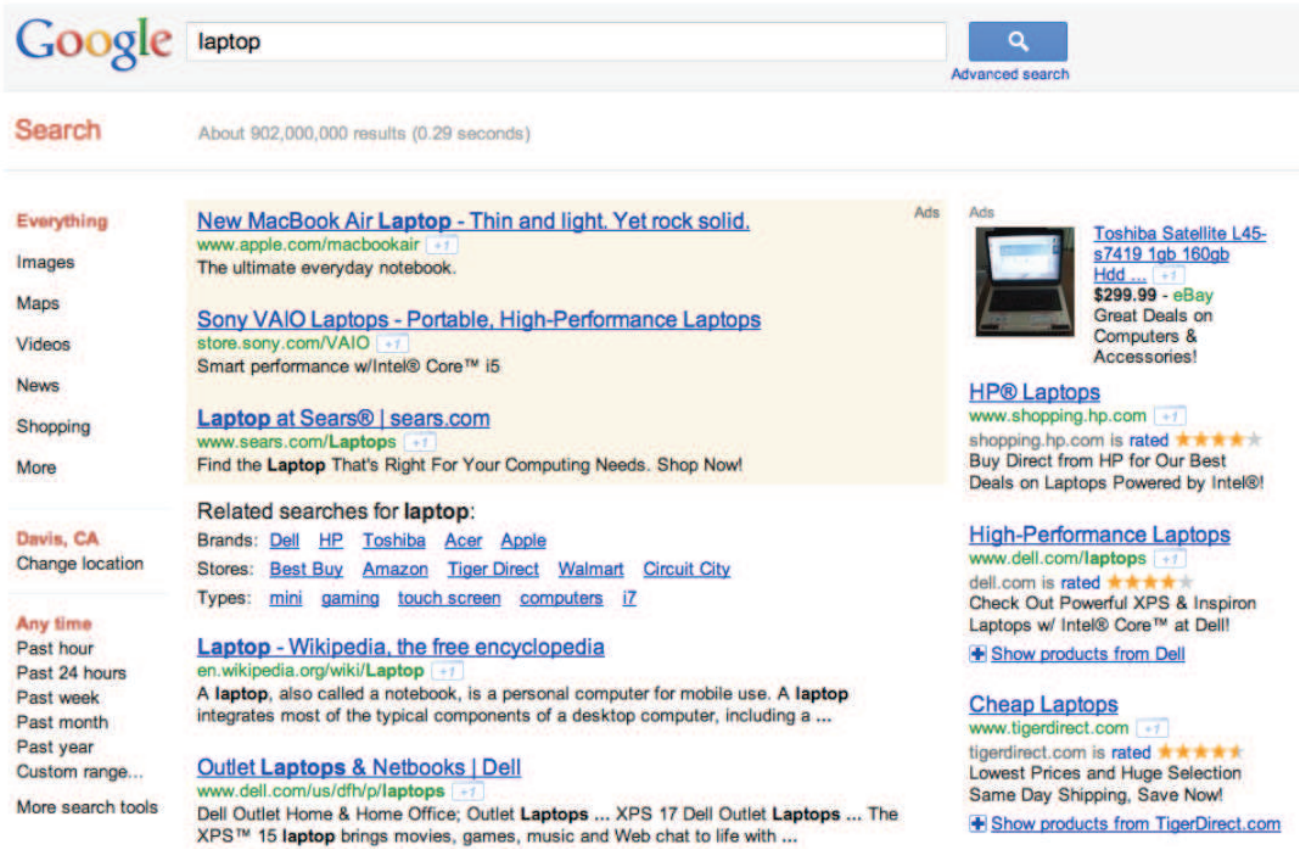
Conversion Effects of Online Advertising Vehicles: Nonparametric Bounds and Parametric Estimates

Shu Shen UC Davis
Lizhen Xu GA Tech
Jianqing Chen UT Dallas
Andrew Whinston UT Austin

November 15, 2011

Different Types of Online Ads

- Sponsored Search
- Display Banner
- Email Ads



Size of the Industry

- The IAB-PWC 2008 Report
  - The total revenue is 23.4 billion in 2008.
  - 45% was paid for sponsored search ads.
  - 21% was paid for banner ads.

How Online Advertising Works

- The Impression Effect
  - Information
  - Public image
- The Click-through Effect: unique to online advertising

Question of the Paper

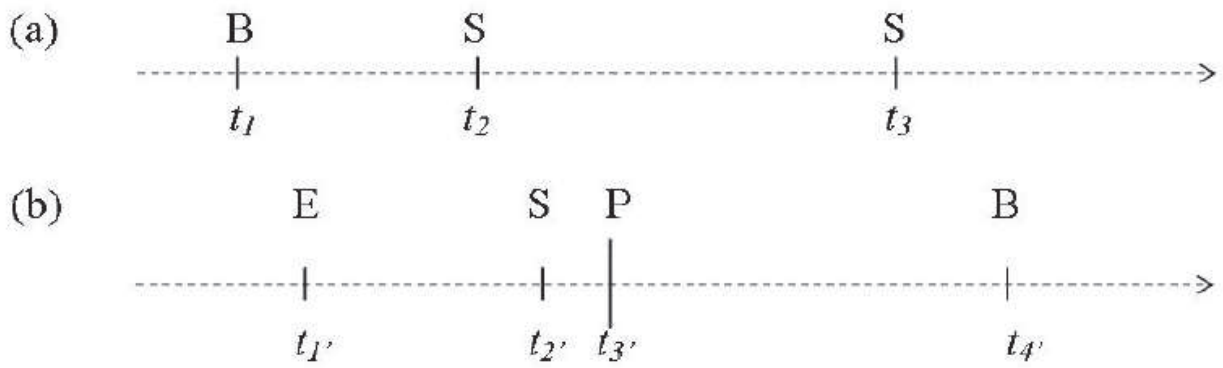
- Question: what is the tangible outcome by which ad clicks are converted to purchases?
  - The conversion rate
  - The conversion dollar

Previous Empirical Literature

- User Browsing Behavior and Future Website Visit Prediction
  - Park and Fader 2004; Montgomery et al. 2004; Johnson et al. 2004
- Conversion Effect of Website Visits
  - Moe and Fader 2004
- Conversion Effect of Banner Ad Exposure
  - Manchanda et al. 2006
- Browsing behaviors are assumed exogeneous even for studies on conversion effects.

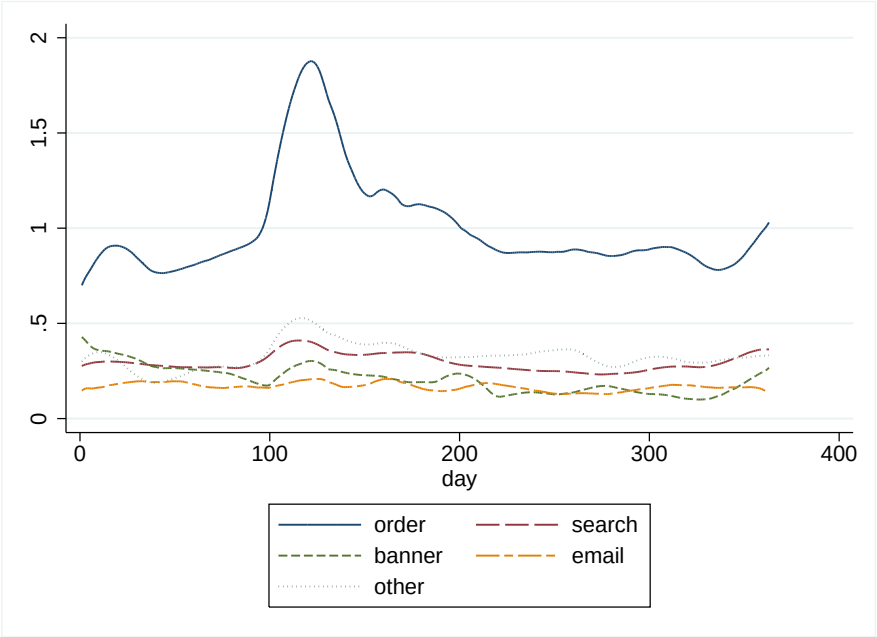
The Full Dataset

- Time Period: August 4th 2007 to July 31th 2008
- Source: a major seller of electronics that sells most of its products online within the time period.
- Consumer Identifier: Cookies
- Consumer enters the dataset through one type of ad-click
- All subsequent ad click and purchase behaviors of the consumer are documented until July 31th, 2008
- Detailed purchase transaction information (product(s) sold, revenue, margin) if purchases are made.



Daily Aggregate Analysis

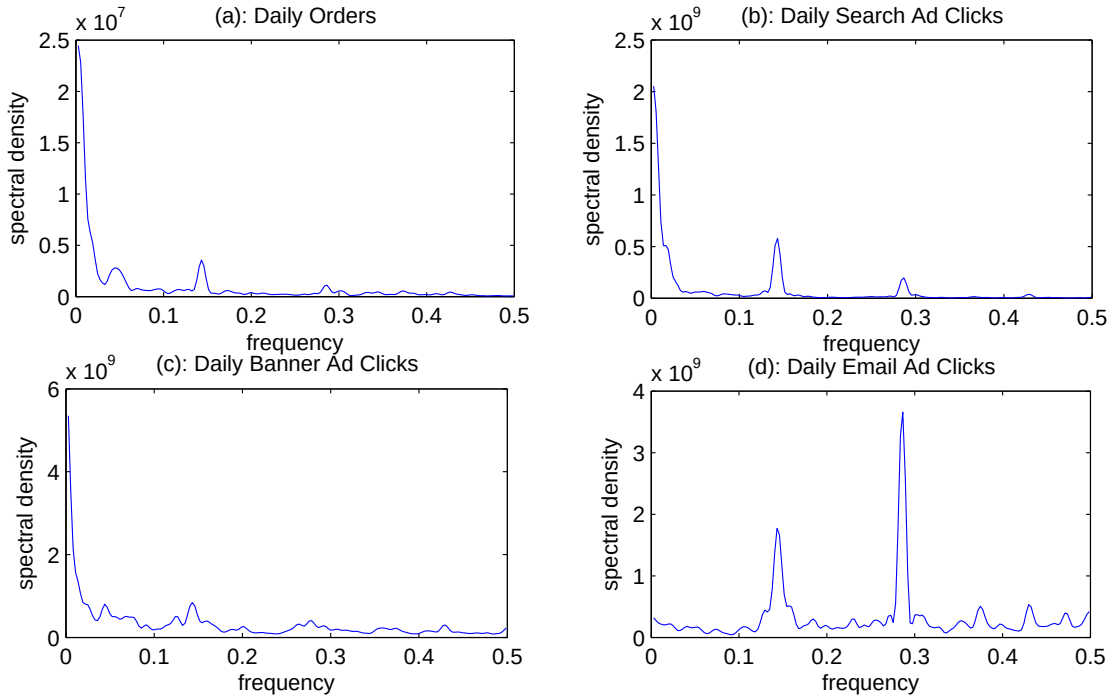
Figure: Purchase and Ad Click Fluctuations During the Year



- Kernel smoothed normalized daily purchases/ad clicks.
- Huge "holiday effect" for daily purchase.
- Small "holiday effect" for daily search and banner ad clicks.

Spectrum Analysis

Figure: Cycles Daily Purchase/Ad Click Numbers



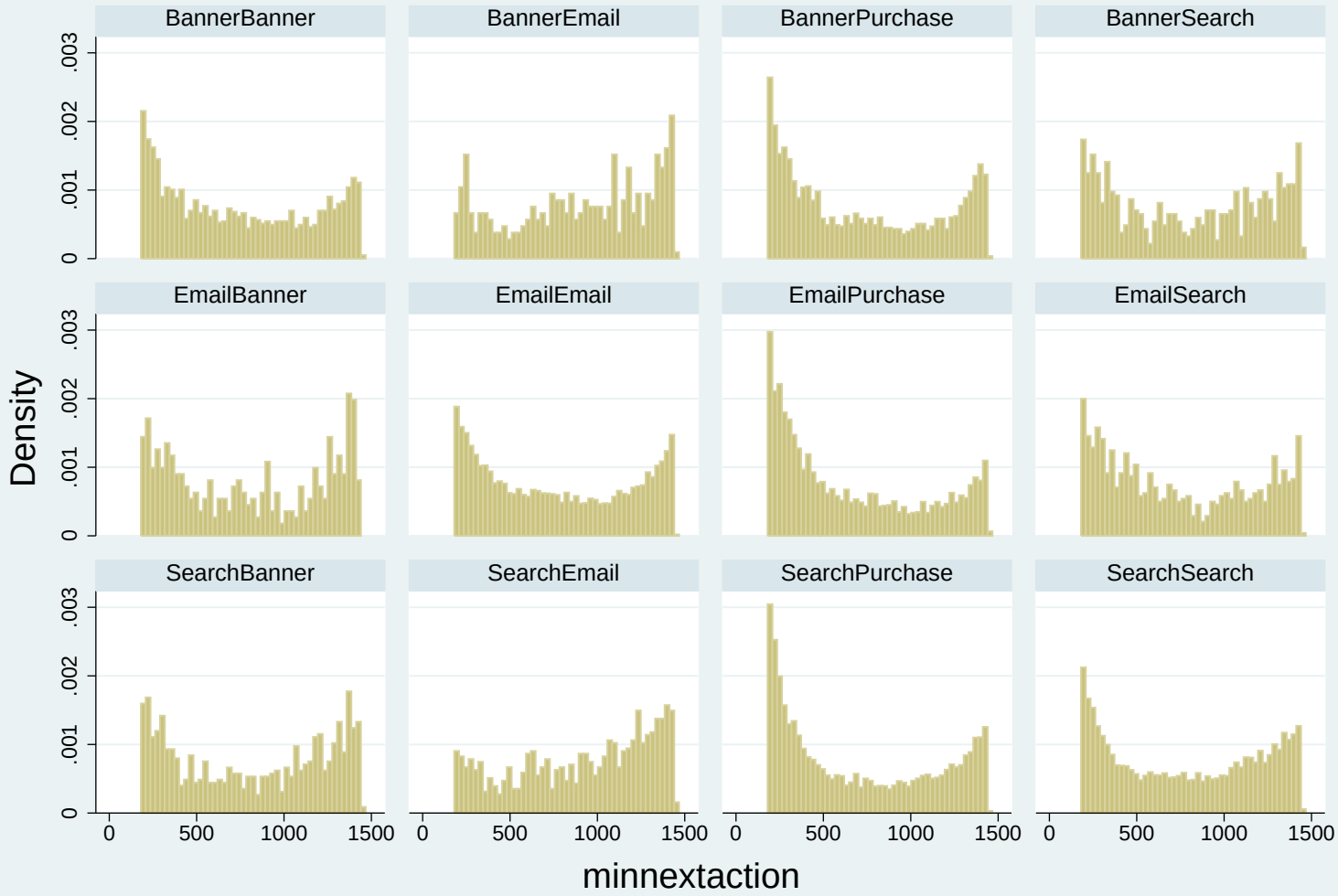
- Big half-weekly cycle of banner ad clicks.
- Weekly cycle of search ad clicks.

The Sample

- Time Period: May and June 2008
- Rules: no activity documented in April 2008 (no left censoring); cannot enter the sample through purchase; no observed other ad click activities; dropped out of the sample once purchase is made.
- Sample Size: around 2 million

Dynamics of Consumer Clickthrough Behaviors

Dynamics of Consumer Clickthrough Behaviors



Dynamics of Consumer Clickthrough Behaviors

The Treatment Effect Framework

- Treatment: T_{is}, T_{ib}, T_{ie}
- Outcomes: $Y_i(t_s, t_b, t_e)$
  - Observed: $Y_i(t_s, t_b, t_e)$ if $T_{is} = t_s,\ T_{ib} = t_b,\ T_{ie} = t_e$
  - Counterfactuals: $Y_i(t_s, t_b, t_e)$ if $T_{is} \neq t_s$ or $T_{ib} \neq t_b$ or $T_e \neq t_e$
- ATE: $E[Y_i(t_s, t_b, t_e) - Y_i(t'_s, t'_b, t'_e)],\ \forall (t_s, t_b, t_e) > (t'_s, t'_b, t'_e)$

Simplified Single Treatment Model

- Drop out individuals who interact with multiple types of ads and group the rest of the individuals into three groups.
- Treatment: T_i
- Outcomes: $Y_i(t)$
  - Observed: $Y_i(t)$ if $T_i = t$
  - Counterfactuals: $Y_i(t)$ if $T_i \neq t$
- ATE: $E[Y_i(t) - Y_i(t')]$, $\forall t > t'$

Data Evidence

Table: The Structure of Ad Click Data

#	S Only	B Only	E Only	S-B	S-E	B-E
1	0.56	0.24	0.20	0	0	0
2	0.53	0.14	0.25	0.06	0.02	0.01
3	0.46	0.10	0.29	0.07	0.05	0.02

- Do we really care about the ATEs for the full sample?

Data Evidence

- Different unobserved purchase incentive might be attached with different ad click behaviors including different types of ad clicks and different numbers of ad clicks.

Table: Evidence of Unobserved Purchase Incentive for Those Purchased

#	Group	Revenue [1]	Cost [1]	M/C If Pos.	If Neg. M [2]
	Search	0.80 (0.98)	0.78 (0.92)	1.59 (1.37)	0.11 (0.32)
1	Banner	0.94 (0.98)	0.95 (0.93)	1.36 (0.92)	0.20 (0.40)
	Email	1.04 (1.00)	1.06 (0.94)	1.27 (0.82)	0.21 (0.41)
	Search	1.06 (0.99)	0.94 (0.97)	1.49 (1.18)	0.13 (0.34)
2	Banner	1.27 (1.06)	1.08 (0.93)	1.30 (0.80)	0.23 (0.42)
	Email	1.18 (0.97)	1.21 (0.92)	1.19 (0.58)	0.25 (0.43)
	Search	1.09 (1.07)	1.08 (0.99)	1.39 (0.99)	0.15 (0.36)
3	Banner	1.15 (1.02)	1.18 (0.93)	1.26 (0.72)	0.24 (0.43)
	Email	1.24 (0.94)	1.29 (0.88)	1.16 (0.48)	0.26 (0.44)

[1] Dollar values are divided by average revenue/cost.

[2] Dummy variable indicating negative margin of the purchase.

The Manski Bounds: MTR+MTS

- Assumptions:
  - MTR: $t_2 \geq t_1 \Rightarrow Y_j(t_2) \geq Y_j(t_1), \forall j = 1, .., N$
  - MTS: $t_2 \geq t_1 \Rightarrow E[Y(t)|T = t_2] \geq E[Y(t)|T = t_1], \forall t$
- A necessary condition:
  $t_2 \geq t_1 \Rightarrow E[Y|T = t_2] \geq E[Y|T = t_1]$
- ATE Bounds:

$$
\begin{aligned}
E[Y(t)] &= E[Y(t)|T = t]P[T = t] + \sum_{s<t} E[Y(t)|T = s]P[T = s] + \sum_{s>t} E[Y(t)|T = s]P[T = s] \\
&\leq E[Y(t)|T = t]P[T = t] + \sum_{s<t} E[Y(t)|T = t]P[T = s] + \sum_{s>t} E[Y(s)|T = s]P[T = s] \\
&\leq E[Y(t)|T = t] \sum_{s\leq t} P[T = s] + \sum_{s>t} E[Y(s)|T = s]P[T = s]
\end{aligned}
$$

The Manski Bounds: MTR+MTS

- ATE Bounds:

$$LB(t) = E[Y(t)|T=t]\sum_{s\geq t} P[T=s] + \sum_{s<t} E[Y(s)|T=s]P[T=s]$$

$$\leq E[Y(t)]$$

$$\leq E[Y(t)|T=t]\sum_{s\leq t} P[T=s] + \sum_{s>t} E[Y(s)|T=s]P[T=s] = UB(t)$$

$$\Rightarrow LB(t_2) - UB(t_1) \leq E[Y(t_2) - Y(t_1)] \leq UB(t_2) - LB(t_1) \quad \forall t_2 \leq t_1$$

- Lower Bound of ATE is always NOT sharp.
- ATE Bounds: $[0\ UB(t_2) - LB(t_1)]$, $\forall t_2 \leq t_1$

Naive Estimates

- $E[Y(t_2)|T = t_2] - E[Y(t_1)|T = t_1]$

Table: Naive Estimates (%)

	Point Estimates			95% CI		
$t_1 < t_2$	S	B	E	Search	Banner	Email
1-2	1.52	0.63	2.51	[1.49 1.64]	[0.59 0.74]	[2.46 2.69]
1-3	2.71	1.71	4.65	[2.62 2.96]	[1.57 2.10]	[4.52 5.03]
1-4	4.59	2.05	7.83	[4.44 4.99]	[1.86 2.58]	[7.63 8.33]
2-3	1.19	1.08	2.13	[1.08 1.45]	[0.94 1.48]	[1.99 2.57]
2-4	3.07	1.42	5.32	[2.91 3.50]	[1.23 1.98]	[5.12 5.86]
3-4	1.88	0.34	3.19	[1.71 2.38]	[0.12 1.00]	[2.95 3.83]

The Manski Bound Estimates: MTR+MTS

Table: Manski Bound Estimates (%)

	Interval Estimates			95% CI		
	Search	Banner	Email	Search	Banner	Email
1-2	[0 1.59]	[0 0.65]	[0 2.74]	[0 1.67]	[0 0.75]	[0 2.88]
1-3	[0 2.74]	[0 1.71]	[0 4.74]	[0 2.94]	[0 2.05]	[0 5.07]
1-4	[0 4.59]	[0 2.05]	[0 7.83]	[0 4.94]	[0 2.56]	[0 8.27]
2-3	[0 2.53]	[0 1.66]	[0 4.26]	[0 2.73]	[0 2.00]	[0 4.61]
2-4	[0 4.38]	[0 2.00]	[0 7.36]	[0 4.73]	[0 2.51]	[0 7.80]
3-4	[0 4.34]	[0 4.68]	[0 7.22]	[0 4.68]	[0 2.48]	[0 7.65]

- The naive estimates always fall in the MTS+MTR bounds derived since the naive estimation imposes ITS which is stronger than MTS.

The Manski Bounds: MTS+MTR+IV

- Assumptions:
  - MTS+MTR
  - IV: $E[Y(t_2) - Y(t_1)] = E[Y(t_2) - Y(t_1)|X = x]$, $\forall t_1 < t_2, x$
- ATE Bounds:

$0 \leq E[Y(t_2) - Y(t_1)] \leq \min_x\{UB(t_2|x) - LB(t_1|x)\}$

The Manski Bounds: MTS+MTR+IV

- Assumptions:
  - MTS+MTR
  - IV: $E[Y(t_2) - Y(t_1)] = E[Y(t_2) - Y(t_1)|X = x]$, $\forall t_1 < t_2, x$
- ATE Bounds:

  $0 \leq E[Y(t_2) - Y(t_1)] \leq \min_x\{UB(t_2|x) - LB(t_1|x)\}$
- X: the weekday (MTWTFSS) when a consumer first enters the dataset
- Informal check of the IV assumption: regress Y on T, X and the interaction between T and X.

The Manski Bound Estimates: MTS+MTR+IV

Table: Manski Bound Estimates (%)

	Interval Estimates			95% CI		
	Search	Banner	Email	Search	Banner	Email
1-2	[0 1.36]	[0 0.48]	[0 1.77]	[0 1.45]	[0 0.54]	[0 1.97]
1-3	[0 1.74]	[0 0.95]	[0 3.13]	[0 2.18]	[0 1.21]	[0 3.55]
1-4	[0 3.96]	[0 1.41]	[0 6.04]	[0 4.26]	[0 1.58]	[0 6.67]
2-3	[0 1.56]	[0 0.89]	[0 2.80]	[0 1.98]	[0 1.16]	[0 3.18]
2-4	[0 3.78]	[0 1.37]	[0 5.70]	[0 4.05]	[0 1.51]	[0 6.30]
3-4	[0 3.72]	[0 1.34]	[0 5.61]	[0 4.01]	[0 1.50]	[0 6.18]

- The naive ATE estimates assuming exogenous treatment are biased upwards.
- The naive ATE estimates for banner ad conversion effect are most severely overestimated.

Parametric Panel Estimation Strategies

- Construct a weekly panel from 5% of the click stream subsample.
- Pros
  - Focus on the effect of most recent ad clicks.
  - Study the interaction between different types of ads.
- Approaches:
  - LPM: Arrelano-Bond
  - DCM: Rivers-Voung

The Linear Probability Model

- Model Setup
  - $Y_{it} = X_{it}\beta + z_{it}\gamma + c_i + u_{it}$
  - X_{it}: Log Ad Click Numbers;
  - z_{it}: Approximation of the "Deal" individual i see at time t
  - $E[u_{it}|X_{is}] = 0, \forall s \leq t$

The Binary Model

- Model Setup
  - $Y_{it}^* = X_{it}\beta + z_{it}\gamma + u_{it}$
  - $Y_{it} = 1(Y_{it}^* > 0)$
  - $X_{it} = X_{is}\Pi + z_{it}\lambda + V_{it}$
  - $(u_{it}, V_{it}) \perp (X_{is}, z_{it}), \forall s < t$
- Empirical Recipe
  - First Step: Run $X_{it} = X_{is}\Pi + z_{it}B + V_{it}$, obtain $\hat{V}_{it}$.
  - Second Step: Run $Y_{it} = 1(X_{it}\beta + z_{it}\gamma + \hat{V}_{it}\eta + u_{it} > 0)$

Results

Table: LPM and DCM Estimates of APEs

	LPM			DCM		[1]
	Pooled	FD	FD-IV	Probit	Tobit	Probit-IV
S	0.028***	0.028***	0.017***	0.007***	0.006***	0.006***
B	0.010***	0.012***	0.008	0.004***	0.003***	0.000
E	0.032***	0.031***	0.018***	0.007***	0.006***	0.011***
R/C	−0.002***	0.003	0.0071*	−0.002***	-0.000	0.003
J			12.16%			6.54%

[1] APEs are reported, evaluated at the mean values of RHS variables.

Results

Table: The Interaction Between Different Types of Ad Clicks

	Probit	Tobit	Probit-IV
Search	0.0076***	0.0066***	0.0057***
Banner	0.0042***	0.0037***	-0.000
Email	0.0067***	0.0058***	0.011***
SearchBanner	-0.0019***	-0.0029***	0.0033 ***
SeachEmail	-0.0026***	-0.0020***	-0.0049***
BannerEmail	-0.0017***	-0.0030***	-0.0022***
R/C	-0.0036***	-0.0029***	0.0022

[1] APEs are evaluated at the mean values of RHS variables.

Conclusion

- Obtain nonparametric bound and parametric point estimates of the conversion effect of different types of online adversising vehicles.
- Endogeneity of consumers browsing behaviors plays a big role in conversion effect estimation, especially for banner ad click through contributed purchases.
- Email ad click-throughs are found to have higher conversion effect than search ad click-throughs than banner ad click-throughs.
- The interaction between sponsored search and display banner ad clickths on average has positive effect on purchase.