Information Signals in Sponsored Search: Evidence from Google's BERT *

Poet Larsen
University of Southern California
poet.larsen@marshall.usc.edu

Davide Proserpio
University of Southern California
proserpi@marshall.usc.edu

April 3, 2025

*We thank Odilon Câmara, Anthony Dukes, Soohyun Kim, Nikhil Malik, Dina Mayzlin, Sean Melessa, Dinesh Puranam, Srinivas Tunuguntula, and participants of the 2022 Conference on Artificial Intelligence, Machine Learning, and Business Analytics, 2023 Stanford Quantitative Marketing Conference, and 2023 Marketing Science Conference for their helpful comments and suggestions. We thank the Institute for Outlier Research in Business and the Marketing Science Institute for generous financial support.

Abstract

We study how improvements to search engine interpretation algorithms and the information signals they generate affect sponsored search markets. We focus on two outcomes: the number of advertisers bidding for a query (i.e., competition) and cost-per-click (CPC). We start by developing a theoretical auction model. We find that as the quality of the interpretation algorithm improves, the number of bidders allocated to auctions generally increases for all queries by improving the platform's ability to identify relevant advertisers more often. Despite this, prices may simultaneously decline. Specifically, we find that shifts in auction prices depend on the prevalence of context in a query. For queries lacking context (e.g., shorter queries), prices generally increase. However, for queries with more contextual information (e.g., longer queries), prices may decrease. This can occur when a platform's new algorithm significantly improves contextual interpretation capabilities, enabling the platform to estimate more precise relevancy scores within the auctions, thereby increasing bidder differentiation and weakening competition among bidders. We then test the model predictions using a monthly dataset of competition scores and CPC for 12,000 queries, leveraging Google's October 2019 rollout of Bidirectional Encoder Representations from Transformers (BERT) as a natural experiment. Employing a Difference-in-Differences identification strategy, we find results consistent with the theoretical model. Our results offer insight into the economic impact of AI and Large Language Models on advertising markets and help advertisers prepare for future interpretation algorithm updates.

1 Introduction

Sponsored search continues to be a dominant advertising channel for firms to reach consumers. It also remains the primary source of revenue for many search engines. In 2023 alone, US search advertising revenue reached nearly $90 billion.[1]

When a consumer submits a search query, a search engine faces the fundamental problem of interpreting the query to generate a response. Search engines such as Google and Bing increasingly rely on Natural Language Processing (NLP) algorithms to interpret consumer search queries and generate search intent signals. These signals help identify and rank relevant advertisers for sponsored search auction opportunities.

At times, search engines update their interpretation algorithms, which can impact the quality of the search intent signal received by the search engine and subsequently influence the sponsored search auction market. These changes are becoming more prevalent due to recent advancements in NLP research. In June 2017, researchers at Google Brain introduced the transformer model architecture (Vaswani et al., 2017). Notable for its ability to dynamically accommodate contextual language information through its "self-attention" mechanism, transformers significantly improved the ability to interpret (encode) and generate (decode) text. This technology has since become the backbone of modern-day Large Language Models (LLMs) and has proclaimed a new era of programmatic interpretation capabilities.

Advertisers are rank-ordered based on ad ranks in sponsored search auctions. Ad ranks depend on submitted bids and relevancy scores. Relevancy scores measure the match quality (i.e., relevance) between an advertiser and a search query and impact both auction eligibility and final paid prices.[2] Higher relevancy scores will boost ad ranks, increasing an advertiser's likelihood of being matched to an auction and winning an ad slot. Higher scores also generally reward advertisers with lower final prices. Query interpretation algorithms provide search

[1]See: `https://www.iab.com/news/2023-u-s-digital-advertising-industry-hits-new-record-according-to-iabs-annual-internet-advertising-revenue-report/`

[2]See `https://support.google.com/google-ads/answer/1722122?hl=en` for discussion of Google ad rank and use of relevancy scores.

intent signals to the seller (Google) to help estimate advertiser relevancy scores. We aim to understand how improvements to the quality of a search engine's interpretation algorithm affect relevancy score estimates and subsequently impact short-term query-level cost-per-click (CPC) and the number of bidders present in auctions (i.e., competition).

The answer to this question is not trivial. In our context, the platform (Google) faces a fundamental trade-off between market thickness and allocation efficiency (Bergemann et al., 2021). On the one hand, new information generated by an improved interpretation algorithm could help the platform better interpret search queries and identify more relevant advertisers, leading to more bidders per auction. If the platform is cautious about allocating irrelevant advertisers, then more informative signals may help it identify relevant bidders more often and lead to potentially higher prices and thicker markets.

On the other hand, new information may help the platform segment markets and remove irrelevant advertisers from auctions, leading to fewer bidders per auction. However, from the platform's perspective, eliminating irrelevant advertisers may not be beneficial, given it lowers the density of the auction market. Instead, it may be worthwhile for the platform to keep less relevant advertisers in the auction, but lower their relevancy score.

The effect on prices (CPC) is ambiguous due to potential changes in the number of bidders competing for a search query, the types of bidders competing, and advertiser relevancy scores. A higher (lower) number of bidders will lead to higher (lower) prices, but increased relevancy scores may lead to lower CPC. If the advertisers that are competing in these auctions are more relevant, i.e., with higher relevancy scores, this may lead to decreasing prices. Advertisers could also adjust bids, but in the short term, they are not receiving any new direct information from the platform, making it unclear how or in which ways they could adjust their bidding strategies.

Closely related to our setting, existing literature on information disclosure in auction markets highlights under what conditions prices may go up or down when buyers receive new information. Information revelation to bidders can lead to better matches, improved allocation efficiency, and higher prices (Milgrom and Weber, 1982; Board, 2009; Tadelis and Zettelmeyer, 2015). But, it can also lead to potentially thinner markets and lower prices (Levin and Milgrom, 2010; Cowgill and Dorobantu, 2020). Whether prices increase or decrease ultimately depends on whether the decline in market thickness lowers the number of competing bidders and drives prices down or allocation efficiency raises bids and drives prices up (Hummel and McAfee, 2016). However, in our setting, the buyers do not observe the interpretation information generated by the language model, meaning the underlying mechanism causing changes to market prices and competition are potentially different.

In this paper, we develop (and test with search ads data) a theoretical auction model to better understand how new information (learned via better algorithms) affects the number of auction competitors and CPC in the short run.[3] Our model considers a platform (e.g., Google) that faces a two-sided matching problem: matching relevant advertisers to incoming consumer search queries. The platform uses an algorithm to learn about the search intent of a query and subsequently uses this information to identify and rank relevant advertisers via relevancy scores. Selected advertisers then compete in the auction for query advertising space. Advertiser relevancy scores impact auction eligibility and final paid prices. Advertisements are sold in a second-price pay-per-click (PPC) auction.

The model analysis focuses on the comparative statics of this partial equilibrium short-run model to understand what immediate impacts the introduction of a new interpretation algorithm may have on the market. Several interesting results stem from the model. First, our model predicts that queries will generally see thicker, not thinner, markets. This translates to more queries with auctions and more bidders within existing auctions. The mechanism behind this result is the negative externality cost the platform incurs if it shows irrelevant ads. More informative signals generated by a new algorithm increase the platform's confidence that it understands the query, which leads to relevant bidders being identified more

[3]The focus on the short run is due to data limitations that do not allow us to study and test long-run effects. Specifically, in March 2020, the COVID-19 pandemic started and likely affected consumer, advertiser, and platform behavior in ways that could directly affect our empirical estimates (e.g., more people spending more time at home and online).

often, driving the average number of bidders up, not down.

Second, the model predicts that the effects on CPC depend on the amount of contextual information (i.e., linguistic modifier information) contained within a query. Context is a distinct type of information that is present in only certain queries.

For queries that contain substantial contextual information (e.g., longer queries such as "running shoes with no laces"), our model predicts that short-run prices may fall despite an increase in the average number of bidders competing for advertising space. This result is driven by the presence of context within a query and an algorithm's ability to understand context. When an algorithm better understands the context in a query, it enables the platform to estimate more precise relevancy scoresand prioritize highly relevant advertisers (i.e., those with high expected click-through rate (CTR)) within the auction. This can lead to softer competition among bidders and lower final prices despite there being a simultaneous increase in the average number of bidders. Our findings complement existing work on information acquisition in auctions and suggest that market allocation efficiency and market thickness might not always be in conflict (Levin and Milgrom, 2010; Cowgill and Dorobantu, 2020). In the context of a matchmaking seller with relevancy scores, auctions can experience thicker markets, lower prices, and improved advertising relevancy.

Queries with limited contextual information (e.g., short queries such as "shoes") will generally see prices and the number of auction competitors increase. The inherent lack of context limits the platform's ability to estimate more precise relevancy scores and further differentiate bidders within auctions.

We rely on a recent event in which Google updated its search interpretation algorithm to test these theoretical predictions. In October 2018, Google researchers used the transformer architecture to build Bidirectional Encoder Representations from Transformers (BERT), one of the first LLMs. BERT was notable for its dramatic improvement at predictive tasks over previous state-of-the-art models (Devlin et al., 2018). In October 2019, Google introduced BERT as one of its main interpretation algorithms to parse and understand user-generated search queries.[4]

To study BERT's effect, we collect monthly average CPC and Competition Score (CS, a measure of the number of bidders participating in an auction) from SEMRush—a company that tracks search query performance—for a sample of roughly 12,000 search queries over two years (2018-2020).

To estimate the effect of BERT on CS and CPC, we exploit the panel nature of our data and employ a difference-in-differences (DD) approach akin to those employed in Eichenbaum et al. (2020), Bollinger et al. (2022), and Liaukonyte et al. (2022). Specifically, we compare changes in CPC and CS before and after the introduction of BERT, with a baseline of changes over the same months and queries but in the previous year. This strategy exploits variation within queries and across time to identify BERT's effect.

To test for market expansion effects, we focus on queries that did not have competition prior to BERT. Among these queries (about 20% in our dataset), roughly 11.5% of them experienced an increase in CS after the introduction of BERT. These results are consistent with our theoretical model and reinforce the idea that new algorithms help the cautious platform expand markets and allocate relevant advertisers to more advertising opportunities.

To test how the amount of contextual information present in a query affects CS and CPC, we group queries based on their length (in general, the amount of contextual information present within queries increases with query length). Consistent with the theoretical model, we find that CS increases by 1.3% for both short and long queries, while CPC increases for short queries by 3.8% and declines by -3.8% for longer queries.

These results withstand several robustness checks, including using additional years as control, and accounting for changes in organic rank results, changes in consumer search behavior, and changes in advertiser behavior. In addition, we replicate these results using

[4]See `https://blog.google/products/search/search-language-understanding-bert/` for the official BERT announcement. See `https://searchengineland.com/welcome-bert-google-artificial-intelligence-for-understanding-search-queries-323976` for industry announcement. See `https://twitter.com/searchliaison/status/1204152378292867074` for the announcement of international BERT stating that US English BERT was introduced in October 2019.

an alternative identification strategy. The idea behind this strategy is to use query linguistic properties to identify queries more or less likely to be affected by BERT. In doing so, we estimate the impact of BERT using an identification strategy similar to a traditional DD with continuous treatment variables.[5].

Combined, our theoretical model and empirical findings help advertisers understand how improved query interpretation algorithms affect sponsored search markets. In particular, our results highlight the role of contextual information present in search queries. For advertisers, the lack of context in short, simple queries inherently limits the ability to differentiate bidders, suggesting that future algorithms will cause these queries to become increasingly more competitive and costly. At the same time, future algorithms will reward advertisers with more precise relevancy score estimates in long-query markets, leading to greater bidder differentiation within auctions. Whether this translates to lower prices depends on the new algorithm's signal improvements.

For academics, our empirical findings contribute to the sponsored search literature and the growing literature on the economic impact of AI and LLMs. Additionally, our theoretical model offers testable predictions for future algorithm updates and improves our theoretical understanding of LLMs' benefits.

2 Related Work

Our paper relates to the growing literature on sponsored search, the economics of AI and LLMs, information disclosure in auction markets, and targeted advertising.

Sponsored Search Sponsored search continues to be a prevalent advertising channel for firms. It also remains an active area of research in both marketing and economics (Edelman et al., 2007; Ghose and Yang, 2009; Yang and Ghose, 2010; Rutz and Bucklin, 2011; Berman and Katona, 2013; Blake et al., 2015; Edelman and Lai, 2016; Simonov et al., 2018; Cowgill

[5]Several papers relied on this type of identification strategy in the past, including **?????**

and Dorobantu, 2020; Sahni and Zhang, 2024). Motivated by the auction structure of sponsored search, Edelman et al. (2007) and Varian (2007) study equilibrium bidding strategies in generalized second price auctions. Empirical work has also analyzed ad effectiveness (Ghose and Yang, 2009; Blake et al., 2015), sponsored and organic complementarities (Yang and Ghose, 2010), keyword spillovers (Rutz and Bucklin, 2011), and advertising competition Simonov et al. (2018). One stream of research focuses on the downstream consequences of search engine platform design decisions, including the impact of a search engine's services on click behavior (Edelman and Lai, 2016), result page features (Gleason et al., 2023), and the interaction between search engine optimization (SEO) and sponsored links Berman and Katona (2013). We contribute to this literature by empirically and theoretically studying how improved search engine interpretation algorithms affect sponsored search markets.

Economic Impact of AI and LLMs Recent advancements in LLM technology have spurred academic interest in understanding the potential economic effect of these models. More recently, the development of ChatGPT has motivated researchers to study how generative LLMs impact areas such as labor markets (Eloundou et al., 2023; Zarifhonarvar, 2023), information markets such as Stack Overflow and Reddit (Burtch et al., 2023), and marketing practices (Kushwaha and Kar, 2021; Reisenbichler et al., 2022; Goli and Singh, 2024). We contribute to this growing literature by studying how LLMs used to generate better signals about consumer search queries impact sponsored search auction markets.

Information Disclosure in Auction Markets Research on information disclosure in auction markets has primarily focused on settings where sellers can endogenously hide or reveal information to buyers. Ganuza (2004) and Board (2009) find that revealing information to buyers about object features when markets are thick generally leads to increasing profits. However, this may not occur in sparse markets due to what Board (2009) calls the *allocation effect*. The allocation effect occurs when information causes the rank ordering of bidder types to swap, leading to weakly decreasing prices.

Related to our setting, Cowgill and Dorobantu (2020) empirically studies how disclosing new information to advertisers in the sponsored search market affects prices, profits, and CTR. The authors find that information disclosure generally leads to thinner markets and lower prices but higher overall profits due to improved CTR. These market adjustments are due to advertisers improving their self-selection into preferred markets, leading to better query-advertiser matching.

Tadelis and Zettelmeyer (2015) also studies how information disclosure affects buyer selection into auction opportunities. In this paper, the authors find that information disclosure in the automobile resale market can help quality-differentiated buyers self-select into preferred auction opportunities, leading to higher market clearance rates and higher profits across all quality types.

When considering seller trade-offs between strategically revealing and hiding information, the theoretical and empirical literature has argued that there is a fundamental trade-off between keeping auctions dense and improving buyer pricing accuracy (Bergemann et al., 2021). Revealing information may help extract value from buyers (Tadelis and Zettelmeyer, 2015), but can also lead to thinner markets (Cowgill and Dorobantu, 2020).

Google faces the same theoretical trade-offs between keeping auctions dense and improving buyer pricing accuracy in our market setting. However, prior literature has focused on settings where buyers receive information (Milgrom and Weber, 1982; Tadelis and Zettelmeyer, 2015; Hummel and McAfee, 2016; Cowgill and Dorobantu, 2020; Ada et al., 2022). In our context, buyers do not receive any information to help with market selection or bid adjustment. Instead, the seller uses the information to pick buyers for auction opportunities and maximize their expected profits. This incentive and market structure may lead to empirical results that differ from previous work.

Targeted Advertising Our paper relates to a stream of literature on matching and targeting technology improvements in advertising markets (Ada et al., 2022). Amaldoss et al.

(2016) studies how improvements to sponsored search broad match technology affect market entry and seller profits. The authors find that better broad match *bidding* algorithms can lower market entry costs and induce greater auction participation, leading to prices.

Empirical and theoretical work has also documented the benefits and market effects of better-targeted advertising technology. Chandra (2009) empirically studies the newspaper market and finds that better targeting can lead to higher prices due to improved advertiser-audience alignment. Athey and Gans (2010) theoretically studies how targeting technology can lead to an increase in the supply of advertising opportunities, potentially putting downward pressure on prices. We contribute to this literature by studying how more informative search query intent signals impact matching in sponsored search markets.

3 Empirical Context

Sponsored Search Advertising When a consumer types a query into a search engine (e.g., "how to cook chicken"), a search engine must interpret the query and decide what domains (URLs) to present on the Search Engine Results Page (SERP). At the top of the SERP, search engines may offer sponsored search links. These links are bid for in a real-time auction before the SERP loads on the consumer's web browser.

An advertiser must create an ad campaign to begin bidding for sponsored search ad space. A sponsored search campaign consists of seven main components: 1) The ad creatives (i.e., what the sponsored search ads look like), 2) The budget of the campaign, 3) The type of individuals the advertiser wants to target (e.g., target Chicago and Los Angeles residents), 4) The keywords the advertiser wants to target, 5) How much the advertiser is willing to bid for each keyword, 6) how to optimize bidding, and 7) The matching strategy. With these ingredients, an advertiser can begin bidding for sponsored search positions.

Sponsored search auctions follow a pay-per-click (PPC) model. Under PPC, advertisers pay only when a consumer clicks on an ad. Because search engines allow multiple sponsored search links to appear at the top of result pages, most rely on a Generalized Second-Price Auction (GSPA) to sell ad space and rank buyers (Edelman et al., 2007; Varian, 2007). Under a GSPA, advertisers are ranked by bids, and the top N winners for N ad slots show advertisements. If an advertisement receives a click, the advertiser pays the minimum price needed to out-price the next highest bidder (usually by $0.01).

In practice, advertisers are ranked with Ad Ranks. Ad Rank scores depend on many factors, including the advertiser's bid and their ad relevancy score. The relevancy score measures how related an advertiser is to a particular query. Higher scores mean the advertiser is more relevant and predicted to receive a click. Ad relevancy matters because it impacts auction eligibility and CPC, with higher relevancy scores increasing the likelihood of auction participation and lowering final paid prices. [6]

The organic rank results are below the sponsored links on the SERP. These are links that the search engine deems relevant to the given search query. Unlike sponsored links, advertisers do not purchase organic links. Instead, the search engine uses an internal ranking system to decide positions. A firm can appear in both sponsored and organic rank slots. For example, even if Nike bids (and wins) the top ad slot in the sponsored search position for the query "women's running shoes", it can still appear in the organic rank results.

BERT Bidirectional Encoder Representations from Transformers (BERT) is a large-scale neural network-based language model developed by Google in 2018. It is a pre-trained model capable of understanding the context and nuances of natural language text, making it a powerful tool for a wide range of NLP tasks (Devlin et al., 2018). Google introduced BERT in October of 2019.

BERT's novelty comes from its ability to understand the context in language objects (Devlin et al., 2018; Tenney et al., 2019b; Rogers et al., 2021).[7] Effectively capturing contextual information was a significant challenge for previous state-of-the-art NLP algorithms such as

[6]See `https://support.google.com/google-ads/answer/1722122?hl=en`.

[7]In simple terms, this means that BERT can differentiate when the word "mouse" refers to a computer mouse or rodent, depending on the words around "mouse".

Word2Vec (Mikolov et al., 2013) and was considered a breakthrough in NLP research. With contextual knowledge, BERT can better understand language semantics, e.g., when "bank" indicates a financial institution or land alongside a river based on the other words in the sentence. It can also lead to a better understanding of complex syntactic language structures, such as adjectives or adverbs, and non-linear relationships between words. To capture this information, BERT generates a vector for each word based on the other words surrounding the focal word, i.e., its context, using the transformer-based self-attention mechanism (Vaswani et al., 2017). As such, the same word can have a different vector representation depending on its surrounding words.

Google introduced BERT in its search engine in October 2019. In its release notes,[8] Google stated that "by applying BERT models to both ranking and featured snippets in Search, we're able to do a much better job helping you find useful information". In other words, Google uses BERT to identify which websites are related to a specific search query (likely by computing similarity scores using vector representations of queries and websites). While Google does not explicitly mention in its press release how this could affect search ads, discussions with current and former employees suggest that signals generated from models like BERT would be readily accessible by the sponsored search team to help rank advertisers.

4 Modeling the Market

We consider a two-sided matching market where a platform uses text to match potentially relevant advertisers to incoming query advertising opportunities. To facilitate matching, the platform relies on an interpretation algorithm to generate signals about the search intent of the search query typed by a consumer. The platform then uses these signals to estimate relevancy scores for advertisers.[9] Using these scores, the platform selects relevant bidders to compete in the auction. Selected bidders submit bids and are ranked within the auction

[8] https://blog.google/products/search/search-language-understanding-bert/

[9] See https://support.google.com/google-ads/answer/1722122?hl=en for discussion of Google ad rank and use of relevancy scores.

based on their ad rank, which is a combination of their bid and relevancy score. The winning ad (that is, the highest ad rank) is displayed and a click is received depending on the true match quality between the winning advertiser and the query. The winning advertiser pays only if a click is received.

The model has two key components. First, we hypothesize that language is inherently multidimensional. More specifically, a query is defined by its topic (i.e., the focus of the query) and its context (i.e., linguistic modifier information). This observation is motivated by an extensive literature across multiple disciplines, including linguistics, computer science, and neuroscience, that commonly models language as multidimensional (Mnih and Hinton, 2008; Jäger and Rogers, 2012; Miyagawa et al., 2013; Coopmans et al., 2023). Second, we assume the platform exclusively receives interpretation signals, while advertisers do not observe this information. This is consistent with advertisers not observing their relevancy scores at bid time or knowing the exact information that is being used to estimate their scores. We highlight this component as it differentiates our empirical setting from other information disclosure in auction settings where bidders are the ones that receive information (Tadelis and Zettelmeyer, 2015; Cowgill and Dorobantu, 2020; Ada et al., 2022).

A query can be deconstructed into the topic(focus) of the query and its modifier information that conveys additional meaning within the topic. For example, the query "running shoes with no laces" focuses on "shoes" and has modifier information in the form of "running" and "with no laces". "Running shoes with laces" has the same focus ("shoes") but conveys different modifier information, particularly "with laces" instead of "with no laces". Both of these queries are categorically similar, but each contains distinct modifier information that conveys different intents within the given topic category. This modifier information is "context".

In NLP, topic modeling is a prevalent area of research. Even prior to BERT, NLP algorithms like Latent Dirichlet Allocation (LDA), Word2Vec (W2V), or Doc2Vec (D2V) organized textual documents into latent "topics". Ultimately, topic modeling focuses on categorizing and clustering similar documents (Blei et al., 2003; Mikolov et al., 2013; Karvelis et al., 2018). When we consider consumer search queries, categorizing queries based on their topics implies the existence of topic differentiation across queries.

Contextual modifier information is distinct from topical information and plays a fundamentally different role within a query. It conveys additional meaning that differentiates queries within a topic category. Although both “running shoes with no laces” and “running shoes with laces” topically relate to “shoes”, they have different types of contextual information that lead to fundamentally different meanings. The context dimension leads to differentiation within a given topic category.

Yet, not all queries explicitly contain context. The short query “shoes” topically relates to “running shoes with no laces” and “running shoes with laces” but lacks additional contextual information to differentiate its search intent within the topic category. Without context, its difficult to further differentiate the query within the givne topic. The inherent difference in information conveyed within the search query fundamentally differentiates shorter queries from longer queries.

To summarize, these examples highlight that queries are multi-dimensional. Queries can be bucketed and differentiated across topics and further differentiated within topics based on context, but only when context is explicitly present. We capture this structure in our model by defining a query as having a topic dimension that defines the focus of the query and a context dimension that defines the modifier information present within the query. A topic and a context dimension will also define an advertiser’s advertisement. The CTR function for a query will depend on the topic and context alignment between the winning advertiser and the query. A platform’s interpretation algorithm generates both topic and context signals that are used to estimate relevancy scores.

While not novel to our setting, it is essential to note that advertisers are horizontally differentiated (e.g., Nike and Geico are relevant to different queries). Additionally, we assume there are negative externality costs paid by the platform when irrelevant advertising is shown.

Irrelevant ads can annoy consumers, increase consumer search costs, or harm the reputation of the platform (Broder et al., 2009; Berman and Katona, 2013; Lu et al., 2017). We now describe our model.

4.1 Setup

Advertiser An advertiser A_i is defined by a topic $\theta_i \in \{0, 1\}$ and a context $z_i \in \{0, 1\}$. We assume four bidders are in the market, one for each combination of topic and context ($< 0, 0 >$, $< 0, 1 >$, $< 1, 0 >$, and $< 1, 1 >$). Advertisers maintain a private value $v_i \sim U[0, 1]$ for clicks.

Query Queries are also defined by a topic $t \in \{0, 1\}$, a context $q \in \{0, 1\}$. θ and z directly map to t and q types for queries, respectively. There are four possible query types ($< 0, 0 >$, $< 0, 1 >$, $< 1, 0 >$, and $< 1, 1 >$). t captures the underlying topic of the query (e.g., is it about "shoes" or "insurance") while q captures contextual modifier information that can differentiate the query type within a topic. Each query has an equally likely probability of being drawn ($\frac{1}{4}$).

Denote the advertiser who wins the advertising opportunity for query Q by A_w, its type by θ_w, and its context by z_w. The CTR for Q depends on the match quality between the winning advertiser and the query. Define $B_t \in [\frac{1}{2}, 1]$ as the relative importance of topic alignment to the query's CTR. We model query CTR in Equation 1.

$$CTR(A, Q) = B_t I_{\{\theta=t\}} + (1 - B_t) I_{\{z=q\}}, \tag{1}$$

B_t measures the relative importance of topic alignment compared to context alignment. For a query where topic alignment matters more relative to context, B_t is large. This will be the case for shorter queries where context is not present. B_t will be smaller for queries where context alignment is meaningful relative to topic alignment. This will be the case for longer queries. Therefore, we assume B_t is decreasing with a query's length. This captures the intuition that for a short query like "shoes", topic alignment matters because context is not present, while for a query like "running shoes with no laces", both topic and context alignment matters to the query CTR.

Platform Our market considers a platform (Google) that sells query ad space to advertisers. Ad space for a query gets sold via a second-price auction (SPA) where the platform endogenously selects who participates in the auction. The advertising space is sold in a pay-per-click (PPC) structure, so the winning advertiser does not pay unless their ad is clicked.

The platform's profit function is:

$$\pi(\hat{A}, Q) = I_{\{Click\}} \frac{\tilde{Adr}}{r_w} - I_{\{\theta_w \neq t\}} C, \tag{2}$$

where $\frac{\tilde{Adr}}{r_w}$ is the second highest Ad Rank in the auction scaled by the relevancy score of the winner bidder, $I_{\{Click\}}$ indicates a click, $I_{\{\theta_w \neq t\}}$ indicates when there is a mismatch between the query topic and winning advertiser topic, and C is the negative externality cost associated with displaying an irrelevant advertisement to the consumer (i.e., wrong topic). $\frac{\tilde{Adr}}{r_w}$ is structured such that winning advertisers pay the minimum bid price needed to win the position (Berman and Katona, 2013; Amaldoss et al., 2015).

Denote the set of chosen advertisers for an auction opportunity by $\hat{A}$. We assume the platform observes advertiser types (θ and z) before selection. However, the platform needs to use an algorithm to interpret and identify the context and topic of a query. When a query Q arrives, the platform uses an algorithm $a \in \mathbb{R}$ to interpret it. For a query Q with topic t and context q, an algorithm a takes Q as input and outputs a signal $\hat{Q}$ with components $\hat{t}$ and $\hat{q}$ to the platform.

For a query Q and algorithm a, $Pr(\hat{t} = t) = \gamma_t(a)$ and $Pr(\hat{q} = q) = \gamma_q(a)$. The platform learns the true value of t with probability $\gamma_t(a)$, and it receives an inconclusive signal with probability $1 - \gamma_t(a)$. Similarly, the platform learns the true value of the context q with probability $\gamma_q(a)$ and it receives an inconclusive signal with probability $1-\gamma_q(a)$. We assume the inconclusive signal is equivalent to the platform's prior: $Pr(t = 1) = Pr(q = 1) = \frac{1}{2}$. Therefore, $\hat{t} \in \{0, \frac{1}{2}, 1\}$ and $\hat{q} \in \{0, \frac{1}{2}, 1\}$. We also assume that $\frac{\gamma_t}{\partial a} > 0$ and $\frac{\gamma_q}{\partial a} > 0$, i.e., a better algorithm always improves the probability of learning the true value of t and q. Finally, we assume advertisers and the platform know $\gamma_t(a)$ and $\gamma_q(a)$.[10]

After receiving the estimates $\hat{t}$ and $\hat{q}$, the platform picks which advertisers, if any, to allocate to the auction opportunity in order to maximize expected profits. Conditional on being allocated to an auction, it is a weakly dominant strategy for advertisers to bid their valuations v_i (the proof is in Appendix A). Advertisers do not pay unless their ad is clicked. We model the expected CPC for a query with advertiser set $\hat{A}$ by $E[CPC_k(\hat{A})] = E[CTR_k] * E[\frac{\tilde{Adr}}{r_w}]$.

Ad ranks determine the order of bidders. We model ad rank for an advertiser i as $Adr_i = b_i r_i(\hat{Q})$, where b_i is the submitted bid and $r_i(\hat{Q})$ is the relevancy score advertiser i receives from the platform given the information set $\hat{Q}$. The relevancy score is modeled as the expected CTR for advertiser A_i given signal $\hat{Q}$: $r_i(\hat{Q}) = Pr(\theta_i = t|\hat{t})B_t + Pr(z_i = q|\hat{q})(1-B_t)$. We assume that the platform knows the CTR coefficient B_t.

Timing The order of the game is as follows. A query Q is randomly drawn by Nature, producing unobservable components q and t. The platform uses algorithm a to generate signals $\hat{q}$ and $\hat{t}$. The platform then estimates relevancy scores for each advertiser and picks some number of auction participants. If the seller runs an auction, chosen advertisers submit bids, the winning advertiser displays an ad, the consumer decides whether to click given the type alignment, and the platform receives a profit. An algorithm a affects profits, CPC, the average number of query bidders, and CTR by changing the seller's q and t signals.

[10]This assumption stems from the fact that Google announces the algorithm update and releases press articles that describe how new algorithms work.

4.2 Results

Our results will focus on auction outcomes at the query level and show how improving the algorithm a can impact CPC and the number of bidders (NOBs) competing for advertising opportunities in the sponsored search auction market. Interestingly, we will show that under reasonable assumptions, query-level CPC may decline despite an increase in the average number of bidders. We now walk through how this can occur. Formal proofs are in Appendix A.

Auction market CPC and the number of bidders adjust with a, depending on the cost of showing topically irrelevant advertisements (C) and the underlying importance of topic alignment (B_t) to a query's CTR. The cost C dictates the platform's willingness to allocate potentially less relevant advertisers to the auction. If C is small, the platform is more willing to allocate less relevant advertisers to auctions because it will lead to denser auctions and place upward pressure on the final paid price. Even if it shows an irrelevant advertisement, the cost of doing so is low and its overall expected profit is better off because the auctions are denser. If C is large, the platform is not willing to allocate less relevant advertisers to auctions for fear they may win and be shown to the consumer. When C is moderate, the average CPC and number of bidders largely depend on the underlying importance of topic alignment (B_t) on the CTR. Lemma 1 describes the technical conditions for C and B_t that lead to different average CPC and number of bidders for queries and Figure 5 in the Appendix visualizes these regions based on B_t and C.

Lemma 1. *The average CPC and NOBs for a query Q depend on C and B_t. Let $J = \frac{40-60B_t+45B_t^2-16B_t^3}{30(2-B_t)^2}$, $X = \frac{9+18B_t-3B_t^2+16B_t^3}{30(1+B_t)^2}$, $Z = J + X - \frac{B(3-B)}{6}$, $P = \frac{-5+20B_t-25B_t^2+15B_t^3-3B_t^4}{6B_t}$, and $L = X + J - P$.*

1. ***Region 1:*** $C \geq \frac{4}{(2-B_t)^2}[\frac{2}{3} - B_t + B_t^2 - \frac{4}{15}B_t^3]$
   - $E[NOB] = 2\gamma_t$
   - $E[CPC] = \frac{\gamma_t}{6}[1 + B_t - \gamma_q(B_t - 1)^2]$

2. ***Region 2:*** $\frac{4+11B_t}{15} \leq C \leq \frac{4}{(2-B_t)^2}[\frac{2}{3} - B_t + B_t^2 - \frac{4}{15}B_t^3]$

   - $E[NOB] = 2\gamma_t + 4\gamma_q - 4\gamma_t\gamma_q$
   - $E[CPC] = \gamma_t\frac{1+B_t}{6} + J\gamma_q - \gamma_t\gamma_q(J + \frac{(B_t-1)^2}{6})$

3. ***Region 3:*** $B_t \leq C \leq \frac{4+11B_t}{15}$

   - $E[NOB] = 4(\gamma_t + \gamma_q) - 6\gamma_t\gamma_q$
   - $E[CPC] = J\gamma_q + X\gamma_t - Z\gamma_t\gamma_q$

4. ***Region 4:*** $\frac{3}{5} \leq C \leq B_t$

   - $E[NOB] = 4(\gamma_t + \gamma_q) - 5\gamma_t\gamma_q$
   - $E[CPC] = X\gamma_t + J\gamma_q - L\gamma_t\gamma_q$

5. ***Region 5:*** $B_t \leq C \leq \frac{3}{5}$

   - $E[NOB] = 4 - 2\gamma_t\gamma_q$
   - $E[CPC] = (J - \frac{3}{10})\gamma_q + (X - \frac{3}{10})\gamma_t - (Z - \frac{3}{10})\gamma_t\gamma_q$

6. ***Region 6:*** $C \leq \frac{3}{5}$ *and* $C \leq B_t$

   - $E[NOB] = 4 - \gamma_t\gamma_q$
   - $E[CPC] = (X - \frac{3}{10})\gamma_t + (J - \frac{3}{10})\gamma_q - (L - \frac{3}{10})\gamma_t\gamma_q$

Region 1 in Lemma 1 describes the average CPC and NOB when C is large, while Regions 5 and 6 dictate average CPC and NOB when C is small. Regions 2 through 4 describe average query CPC and NOB when C is moderate. When $C > \frac{8}{5}$, all queries live in Region 1. For the rest of the Results discussion, we will zoom in on Region 1 and assume $C > \frac{8}{5}$. See Appendix A.4 for exploration of the other regions.

In Region 1, the platform is extremely conservative about allocating advertisers to auctions because the cost of showing topically irrelevant advertisements is high. Only when the platform generates an informative signal about the query's topic is the platform willing to run an auction with advertisers. This leads to an average number of bidders for a query Q of $2\gamma_t(a)$. Intuitively, this means the platform will only allocate advertisers when they know they're in the right ballpark of the query's focus. If the platform knows an incoming query is about "shoes", it is happy to allocate "shoe" advertisers. But, if the platform is not unsure whether a query is about "shoes" or "insurance", it's not willing to allocate any bidders for fear of incorrectly showing an "insurance" advertiser on a "shoe" query, or vice versa.

When the platform improves the quality of its interpretation algorithm (higher a), this improves its ability to generate informative topic signals. In turn, it improves the platform's ability to categorize queries and ultimately allocate relevant advertisers to relevant auctions. This leads to Proposition 1.

Proposition 1 (Platform Selection of Advertisers)**.** *Under sufficiently high cost C ($C \geq \frac{8}{5}$), the average number of bidders for query ad space increases with a ($\frac{\partial}{\partial a} 2\gamma_t(a) > 0$).*

Average CPC for a query in Region 1 depends on both γ_t and γ_q. Topic signals (γ_t) help the platform get relevant advertisers into the auctions, while context signals (γ_q) help the platform estimate more precise relevancy scores and identify how relevant an advertiser is to a given query when context is explicitly present within the query. When the context dimension is understood with the topic dimension, the platform can do a better job prioritizing the highly relevant advertiser within the auction by increasing their relevancy score and decreasing the relevancy score of the less relevant advertisers. However, estimating more precise relevancy scores also softens the competition between the bidders and can lead to lower CPC. This can occur in query markets where context is explicitly present and matters to the underlying CTR (small B_t). If the platform improves a and significantly improves the precision of the context signals, this competition softening can potentially lead to lower average CPC. This leads us to Proposition 2.

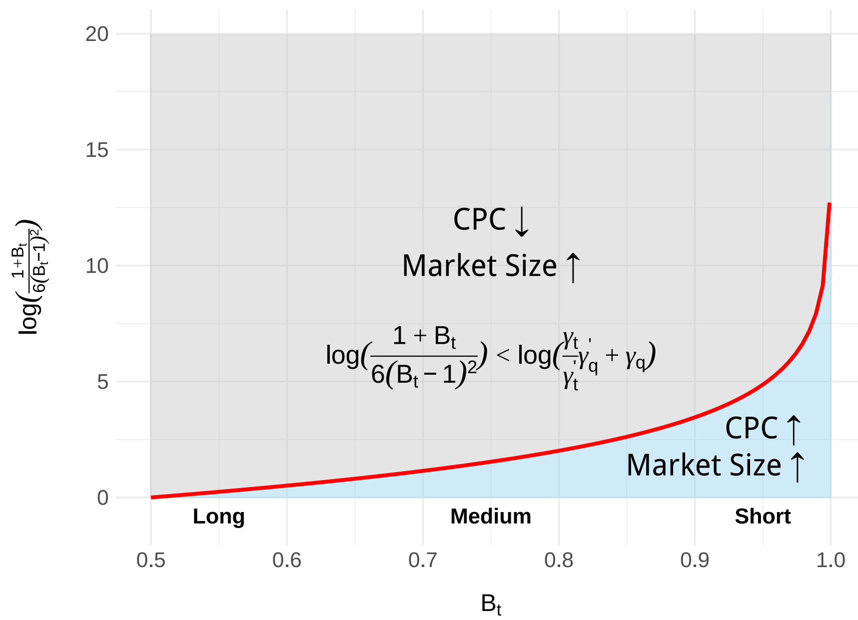
Proposition 2 (Average CPC)**.** *Average CPC for a query in Region 1 will decline if $\frac{1+B_t}{6(B_t-1)^2} < \frac{\gamma_t}{\gamma_t'}\gamma_q' + \gamma_q$, where γ_q' and γ_t' are the partial derivatives of each with respect to a.*

CPC is affected by two forces: better matching due to better topic identification and better bidder differentiation within auctions due to better context identification. Topical understanding improves categorical matching, increasing the average number of bidders and CPC for a given query. Simultaneously, better context signal helps the platform estimate more precise relevancy scores within the auction and further differentiate the topically relevant bidders. Contextual differentiation prioritizes advertisers who are more likely to receive a click and rewards them with lower CPC. This softens pressure on CPC caused by improved topical understanding and potentially enables CPC to decrease. Whether CPC decreases or increases after the introduction of a new algorithm depends on the importance of context alignment for a query's CTR and the relative improvements in the context signal ($\gamma_q^{'}$) the platform receives. If contextual advertiser-query alignment matters a lot to the CTR of a query (as is the case for longer queries where B_t is small), and the contextual signal improvements obtained with a new algorithm are large, CPC may decline. However, if contextual congruency is not important for the CTR of a query (as is the case for short queries where B_t is large), then it is unlikely that CPC declines.

In Figure 1 we present $log(\frac{1+B_t}{6(B_t-1)^2})$ by B_t. We log scale the inequality for visualization and indicate what regions of B_t short, medium, and long query lengths would live. The gray region of the figure indicates when the inequality would hold (i.e., $log(\frac{1+B_t}{6(B_t-1)^2}) < log(\frac{\gamma_q^{'}}{\gamma_t^{'}}\gamma_t + \gamma_q))$ and CPC would decline despite an average increase in the number of bidders. The blue region indicates where the inequality would not hold and CPC would increase with the rise in the number of bidders. The line is where there is equality and there would be no changes to CPC. As B_t decreases, the potential for lower CPC increases. In other words, longer queries have greater potential to see declining CPC.

Summary and Predictions The results of our stylized model are driven by the notion that context is a unique construct that, when understood by the platform, enables the platform to estimate more precise relevancy scores. However, context is not present in all

Figure 1: Potential market changes in Region 1.



queries, and this impacts how the auction market for a query can change in the presence of a new interpretation algorithm.

Given the results of our model, we predict that BERT will lead to expanded markets. This means that queries that were not competitive prior to the introduction of BERT will see an increase in the number of bidders. It also means that queries with existing auctions will, on average, see more bidders. Simultaneously, we expect CPC to adjust for queries with auctions that existed prior to the introduction of BERT based on the amount of contextual information present within the query. We expect that for shorter queries (high B_t) CPC will increase, while for longer queries (lower B_t) CPC will decrease.

5 Empirical Analysis

5.1 Data

Sampling search queries To empirically test the predictions of our model, we collect data from SEMRush, a leading provider of sponsored search rank and keyword data. We rely on a survey that we administered to Amazon Mturkers to generate a sample of queries for the analyses presented in this paper. After selecting 32 different topics (see Appendix B for the list of topics), we asked survey respondents the following question: "Please write a search query related to the topic of '**Topic**' that you would search for on Google". 100 Amazon Mturk participants took the survey and each participant was asked about five randomly selected categories, giving us 500 responses and roughly 16 responses per topic category.

Not all MTurk submissions were queries. To focus on submissions that were search queries, we removed 210 irrelevant answers. Irrelevant responses included answers where participants typed in specific URLs, website content, descriptions of how to type up a search query, and quotes from Google "Help Pages" describing how to search. Removing these responses left us with 290 search query answers.

Using the remaining 290 sampled queries, we turned to the broad match database at SEMRush.[11] For a given query, SEMRush returns a set of similar queries for which it has data for the current month.[12] We limited our match to queries with an average search volume of at least 500 from 2021-2022, generating a dataset of roughly $120,000$ queries related to the original 290 queries. However, many of the queries contained explicit terms (e.g., adult content such as pornographic search queries) unsuitable for analysis (advertisers generally don't target adult content queries). To filter these out, we generated a list of adult content terms and fuzzy-matched each query with the list of terms. We removed the query from our data if there was a reasonable match between the query and any of the terms. After completing this process, we retained roughly $40,000$ queries.

CPC and Competition Score SEMRush provides two outcomes that are relevant to answering our research question: CPC and Competition Score. CPC measures the average price advertisers pay for an ad click. Competition Score is a proprietary normalized score (between 0 and 1) that measures the relative number of advertisers bidding for ad space for

[11] See: `https://www.semrush.com/features/keyword-research/`

[12] For example, for the query "shoes", the database will show monthly data for "shoes" as well as similar queries, such as "running shoes", "women's shoes", "best shoes for hiking", etc.

the given query.[13] It is worth noting that SEMRush collects data for many queries, even if they have little to no advertising, meaning CPC and Competition Score can take on the value of zero.

We collected historical monthly information about CPC and Competition Score for each of these queries from January 2018 to February 2020. We then filtered out queries that did not have persistent historical information (more than a year of missing data) and those with limited historical search volumes (less than 100 average searches/month during the time frame).

To empirically validate Competition Score and CPC, we collect SEMRush and Google Keyword Planner (GKP) tool data from July 2021 to February 2022 for 10,000 queries in our dataset.[14] From GKP we have two measures of competition: (1) “Competition”, which is a measure of the number of bidders competing in the query’s auction using three values: low, mid, and high; and (2) “Competition (index value)” which measures the total number of ad slots filled divided by the total number of ad slots available on a scale from 0 to 100. The correlation between these two measures is 0.98. We then compute the Pearson correlation between the Competition Score of SEMRush and the two competition measures from Google. We find that the Competition Score of SEMRush has a correlation of 0.65 with the Competition measure obtained from Google and a correlation of 0.87 with the Competition (index value) obtained from Google [15], suggesting SEMRush data is reasonably correlated with data obtained directly from Google.

To validate CPC, we use we use the only CPC-related measure from GKP: the maximum submitted bid. While average CPC and maximum bid are different measures, they should reasonably correlate if SEMRush data is reliable. The Spearman correlation between the average CPC of SEMRush and the maximum submitted bid from Google is 0.57, again

[13]From discussions with SEMRush, Competition Score is linearly comparable and can be viewed similarly to the Google Trends information commonly used in research to measure search volume. While we can’t estimate the precise number of bidders, we can estimate relative changes.

[14]Unfortunately, GKP data goes back in time to at most three years, and we did not collect data in time to replicate some of the analysis reported in this paper with GKP data

[15]We encode Google’s Competition measure as 0, 0.5, and 1 to calculate this correlation.

suggesting a high level of agreement.

Our final dataset contains $11,949$ unique queries and $284,146$ monthly observations. In the top part of Table 1, we present overall summary statistics for our dependent variables, Competition Score and CPC, across all queries from January 2018 to February 2020.

Table 1: Summary statistics.

	Mean	St. Dev.	Min	Max	Median
	All queries				
CPC	1.25	3.343	0.00	584.73	0.54
Competition score	0.244	0.356	0.00	1.00	0.05
	Short queries				
CPC	0.911	1.918	0.00	64.110	0.400
Competition Score	0.248	0.364	0.00	1.00	0.050
	Medium queries				
CPC	1.339	3.947	0.00	584.730	0.610
Competition Score	0.278	0.378	0.00	1.00	0.060
	Long queries				
CPC	1.533	3.078	0.00	83.860	0.720
Competition Score	0.126	0.216	0.00	1.00	0.030

Table 1 show that, in the aggregate, 1) many queries are not competitive (low median and mean Competition Score) and 2) CPC is often relatively low, but the maximum price can be extremely high. We also break down queries by word length, an important moderator we focus on in our analysis. We group queries into three categories for ease of presenting results. A short query has two or fewer words, a medium query has three to five words, and a long query has six or more words. We present summary statistics by query length in the bottom part of Table 1. The table shows that as query length increases, demand generally decreases while CPC increases. These results generally align with the trend that, as query length increases, the queries become more niche and valuable to advertisers but also harder to target effectively.

5.2 Empirical Strategy

To study how BERT affects CS and CPC, we implement two identification strategies that rely on different underlying assumptions. In this section, we describe our preferred identification strategy. We discuss the second strategy in Appendix C. Results are consistent across both methods.

Our main identification strategy exploits the panel nature of our data and compares changes in each query's CS and CPC before and after the introduction of BERT with changes in the same query's CS and CPC in the year before the introduction of BERT.[16] In doing so, we implement a strategy akin to a difference-in-differences (DD) where the treated and control queries are the same, but outcomes are observed across years.[17] Our identification strategy is similar to those employed in recent papers, such as Eichenbaum et al. (2020), Bollinger et al. (2022), and Liaukonyte et al. (2022).

This strategy aims to use variation across time within queries to identify BERT. An advantage of this strategy is that selection into the treatment is not an issue since treated and control queries are the same. However, time-varying unobservables that vary at the year-month level may bias our results. For example, our results could be positively biased if some market factor correlated with CS and/or CPF motivated Google to launch BERT in October 2019. Alternatively, queries may not be comparable across years if a query correlates with year-month-specific unobservable events, e.g., large concerts or sporting events. Both issues are classic DD challenges akin to unit-time-specific events tampering with causal effect estimates.

We operationalize this identification strategy using the following model:

$$Y_{ijt} = \beta_1 Treated_{ij} \times Post_t + \delta_i + \gamma_j + \psi_t + \epsilon_{ijt}, \tag{3}$$

[16] Essentially, we are comparing outcomes for each query in the year in which BERT was introduced with outcomes in the year prior to BERT introduction.

[17] In standard DD settings, outcomes are observed over the same period, but treated and control units are different.

where $Y_{i,j,t}$ is the outcome of interests (either $log(CPC)ijt$ or $log(CS)_{ijt}$) for query i, year group j and month t.[18] $Treated_{ij}$ is a binary indicator that takes on the value of one when query i is observed during the treated year-group window (July 2019–February 2020), zero otherwise (July 2018–February 2019).[19] $Post_t$ is a dummy that takes on the value one if the month t is after the month Google introduced BERT (November–February), zero otherwise (July–October). We include query fixed effects (δ_i) to control for time-invariant query unobservables, year-group fixed effects (γ_j) to control for group-specific shocks impacting all queries in the same year-group (e.g., in 2019–20, demand is higher for search ads), and month-of-year fixed effects (ψ_t) to account for monthly shocks common to all queries (e.g., holidays and advertiser monthly spend). We cluster standard errors at the query level to account for potential serial correlations in the dependent variable and estimate the model using OLS. The parameter of interest is β_1, representing the incremental impact of BERT on the sponsored search auction outcomes of interest.

We estimate Equation 3 from July to February so that we have four pre-BERT months (July–October) and four post-BERT months (November–February).[20] In addition, we focus on two periods, July 2018–February 2019 (control period) and July 2019–February 2020 (treated period). In Section 5.4, we show that results hold with additional control year groups, specifically, July 2016–February 2017 and July 2017–February 2018 months.

Identification Checks The key assumption behind our DD identification strategy is that no unobserved time-variant, group-specific shocks correlate with the entry of BERT and auction market outcomes.

In our setting, we worry about changes in advertiser or consumer search behavior due to time-varying unobservable events that correlate with BERT's release. It is possible that

[18]We add 1 to both CPC and CS to avoid taking the log of zero.

[19]We refer to j as a group rather than a year because our treated (control) periods fall across multiple years. For example, the treated period includes the years 2019 and 2020 because the treated period spans July 2019 to February 2020.

[20]We limit the post-BERT period to February because we want to avoid picking up any COVID-19 effect, which started impacting the US in March 2020.

consumer search demand characteristics differ across years and that consumer behavior or advertiser behavior in, say, the holiday season in 2018 is different than that in 2019. If this is the case, the results we observe may be due to these unobservable time-varying factors. This concern is analogous to worrying about unit-specific time-varying unobservable events correlated with the treatment event in a DD across states or cities.

To reduce concerns about these issues, we do two things. First, to alleviate demand-side consumer shocks, our sampling procedure removes queries that have intermittent or volatile search volumes that may correlate with unobservable time-varying events. For example, a query about a political event that may see a spike in search in a particular month and then quickly disappear will not be part of our sample. Doing this restricts our analysis and results to primarily persistently searched queries in order to minimize exposure to volatile, time-varying search queries.

Second, as is common in DD analyses, we show that treated and control units behaved similarly before the introduction of BERT (parallel trends assumption), suggesting that the year prior to BERT is a good counterfactual. We do so by implementing an event study design.

To obtain event study parameter estimates, we estimate the following specifications:

$$Y_{ijt} = \beta_1 Treated_{ij} \times Month_t + \delta_i + \gamma_j + \psi_t + \epsilon_{ijt}, \tag{4}$$

Everything is as in Equation 3, but we replace our binary *Post* variable with eight monthly dummies that indicate the eight-month window from July to February. We estimate Equation 4 for all the outcomes we study, setting the baseline level for the monthly dummies to be October (the month prior to the introduction of BERT). In the next section, we present these event studies along with the main estimates.

5.3 Results

Topical information effect The theoretical model predicts that the topical signal gains due to BERT help Google identify relevant bidders more often for already competitive queries and turning non-competitive queries into competitive ones. To test this prediction, we start by splitting the data into competitive and non-competitive queries. We deem a query non-competitive if the mode Competition Score during the July-October pre-period in 2018 and 2019 is zero. We find that roughly 20% of our queries are deemed non-competitive.

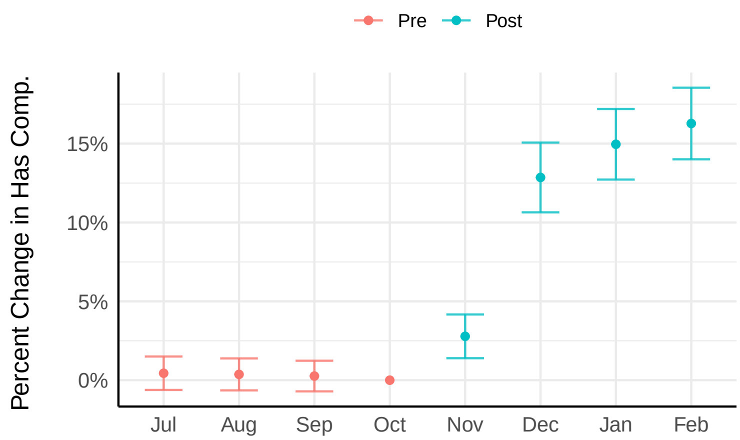
We then take these queries deemed non-competitive and create a binary variable called $Has\ Competition_{ijt}$ that takes on the value one if query i's Competition Score is greater than zero for a given year-group j month t, zero otherwise. Using this variable, we can measure what proportion of non-competitive auctions become competitive in the post-BERT periods. We then estimate Equations 3 and 4, using $Has\ Competition_{ijt}$ as the dependent variable. In Figure 2, we visualize the event study parameter estimates. We find that prior to the treatment, estimates are indistinguishable from zero, with no apparent trends. In the post-treatment period, we observe that the estimates become positive, an effect consistent with BERT increasing ad space supply by identifying more relevant bidders for auction opportunities.

We present the estimates of Equation 3 in Table 2. We find that BERT converts roughly 11.5% of non-competitive queries into competitive markets.[21] This suggests that topical signal improvements due to BERT help the platform identify meaningfully relevant advertisers for more query auctions.

Next, we turn to the sample of competitive queries. Since our theoretical model predicts an increase in competition across all queries (i.e., with more or less contextual information), we group queries by query length. Query length correlates with the degree of contextual

[21] As a sanity check, we estimated Equation 3 using $log(CPC)$ as the dependent variable for the queries deemed non-competitive before BERT. We find a positive and statistically significant coefficient (0.03 with $p = 0.0005$), suggesting that CPC also rises with the increase in the number of competitive auctions for the non-competitive queries following the introduction of BERT.

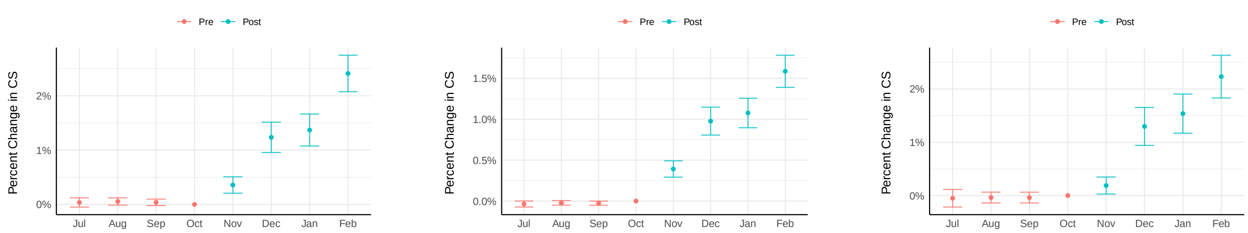
Figure 2: The effect of BERT on Has Competition for non-competitive queries. Error bars represent 95% confidence intervals.



information present within the query and is often used as a rule-of-thumb measure by advertisers when discussing targeting strategies. We then estimate Equation 3 and Equation 4 by query length and using $log(CS)$ as the dependent variable.

We start by presenting the event study estimates in Figure 3. Before the treatment, we find estimates close to zero, partially validating the parallel trends assumption. In the post-treatment period, estimates become positive and significant for short, medium, and long queries, suggesting that competition increases for all already competitive queries.

Figure 3: Percent change in Competition Score by query length. Error bars represent 95% confidence intervals.



(a) Short queries (b) Medium queries (c) Long queries

In Table 3, we report the main effect estimates from Equation 3. We find CS increases between 1% and 1.3% for all queries.

Table 2: The effect of BERT on Has Competition for non-competitive queries.

	(1)
Post × Treated	0.114***
	(0.009)
Observations	30,474
R^2	0.380

Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: Regressions include year-group, month, and query fixed effects. Standard errors clustered at the query level are reported in parentheses.

Table 3: The effect of BERT on $log(CS)$ by query length.

	(1) Short	(2) Medium	(3) Long
Post × Treated	0.013***	0.010***	0.013***
	(0.001)	(0.001)	(0.001)
Observations	47,912	88,888	23,887
R^2	0.951	0.971	0.920

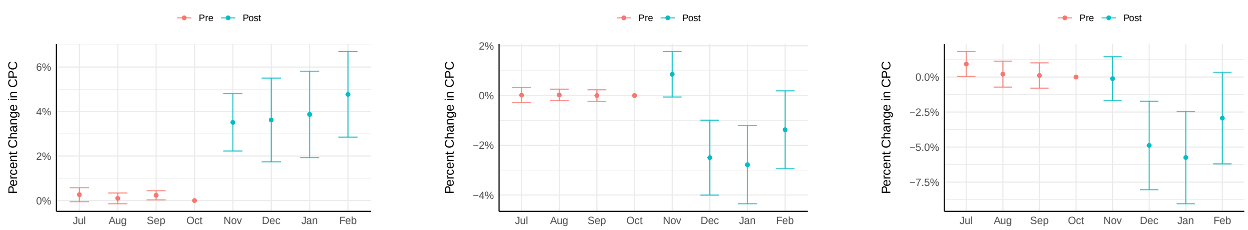
Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: Estimated with competitive queries. Regressions include year-group, month, and query fixed effects. Standard errors clustered at the query level are reported in parentheses.

Consistent with our model's predictions, the empirical evidence suggests that BERT's introduction helped Google uniformly expand auction markets and improve advertiser allocation to auctions. This is because BERT helped Google overcome the negative cost C and do a better job identifying topically relevant advertisers for queries.

Contextual information effects Our model predicts that CPC can increase or decrease depending on whether the query contains contextual information, despite the average number of bidders increasing across all queries. In particular, CPC should increase for queries with limited to no contextual information; however, CPC may decrease for queries with contextual information.

Again, we group queries based on their length to operationalize and test these hypotheses. We then estimate Equation 3 and Equation 4 on the set of competitive queries, by query length, and using $log(CPC)$ as the dependent variable.[22] In Figure 4, we present the event study estimates by query length. First, as before, in all cases, we find parallel pre-treatment trends, supporting the validity of our identification strategy. Second, we see that in the post-treatment period, CPC increases for short queries but decreases for medium and long queries.

Figure 4: Percent change in CPC by query length. Error bars represent 95% confidence intervals.



(a) Short queries. (b) Medium queries. (c) Long query.

In Tables 4, we present the estimates of Equation 3 by query length. Consistent with our model predictions, query length moderates how BERT's introduction affects existing auction market CPC. We find a 3.8% increase in CPC for short queries. However, as query length increases, CPC quickly decreases. CPC decreases by 1.5% for medium queries and 3.7% for long queries. These results imply that BERT's introduction significantly improves Google's ability to estimate more precise relevancy scores for queries that contain contextual information, leading to lower CPC but ultimately more relevant ads.

[22] We again remove non-competitive queries because, for these queries, prices can only increase since they were zero when they were non-competitive.

Table 4: The effect of BERT on $log(CPC)$ by query length.

	(1) Short	(2) Medium	(3) Long
Post × Treated	0.038^{***}	-0.015^{**}	-0.038^{***}
	(0.008)	(0.006)	(0.013)
Observations	47,912	88,888	23,887
R^2	0.686	0.728	0.671

Significance Levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.
Note: Estimated with competitive queries. Regressions include year-group, month, and query fixed effects. Standard errors clustered at the query level are reported in parentheses.

5.4 Robustness Checks

In this section, we discuss three tests aimed at reinforcing our main results. Specifically, we show that our results are robust to the inclusion of additional years as controls, changes in organic search ranking, and changes in consumer and advertiser search behavior. In Appendix C, we also present an alternative identification strategy aimed at identifying BERT's effect using variation in query linguistic properties. This strategy relaxes the year-month unconfoundness assumption made with our primary identification strategy. We find results consistent with our primary identification. For the sake of brevity, we focus on replicating the effects for competitive queries, i.e., an increase in CS across all queries and heterogeneous adjustments to prices depending on the length of the query.

Additional Years as Controls Our primary DD analysis uses query data from 2018 to 2020, with July 2018 to February 2019 as control units for July 2019 to February 2020. The downside of using only one year-month set (2018-2019) as a control window is we don't know whether it accurately represents the behavior of the auction market during those months. (This is equivalent to worrying about using only two states in a traditional state-level DD). As a robustness check, we re-estimate Equations 3 and 4 using 2016-2017 and 2017-2018

July-February months as additional control groups.

To identify competitive queries with this additional data, we take the mode Competition Score across all July–October months and years (2016–2019, generating 16 observations per query) and deem a query non-competitive if the mode is 0. In Figure **??**, we present event analysis market changes for $log(CS)$ and $log(CPC)$ for queries deemed competitive before BERT.

We repeat our regressions by query length for $log(CS)$ and $log(CPC)$. We report these results in Tables 5 and 6. The results are directionally consistent and similar in magnitude to those estimated in Tables 3 and 4. The fact that using additional control years leads to broadly similar results suggests we are reasonably controlling for within-unit seasonal patterns and that the 2018-2019 auction market outcomes are fairly representative of the average Google market in the absence of BERT.

Table 5: The effect of BERT on $log(CS)$ by query length.

	(1)	(2)	(3)
	Short	Medium	Long
Post × Treated	0.012***	0.009***	0.017***
	(0.001)	(0.001)	(0.001)
Observations	83,765	154,872	36,927
R^2	0.917	0.956	0.905

Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: Estimated with competitive queries. Regressions include year-group, month, and query fixed effects. Standard errors clustered at the query level are reported in parentheses. Data from 2016 to 2020.

Changes in Organic Rank Results One potential explanation for our results could be that advertisers respond to changes in the organic ranking results or to new features added to the results page that are a function of BERT. For example, in its release note, Google mentions that BERT will help with featured snippets. Adding more featured snippets at

Table 6: The effect of BERT on $log(CPC)$ by query length.

	(1) Short	(2) Medium	(3) Long
Post × Treated	0.032^{***}	-0.021^{***}	-0.058^{***}
	(0.009)	(0.006)	(0.013)
Observations	83,765	154,872	36,927
R^2	0.629	0.702	0.627

Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: Estimated with competitive queries. Regressions include year-group, month, and query fixed effects. Standard errors clustered at the query level are reported in parentheses. Data from 2016 to 2020.

the top of the page may take away premium space and increase demand for sponsored search. This could cause an increase in the number of bidders and, potentially, prices. Moreover, if the top organic result of a search query changes, this may suggest that BERT has substantially altered the interpretation and identified a new website as more relevant to a given query. A switch or change in the top URL could subsequently impact prices and demand for advertising space.

To address this concern, we collect monthly organic rankings data from SEMRush for the queries in our data. Using this data, we test for changes in the likelihood of a query resulting in a featured snippet, the number of Search Engine Results Pages (SERP) features shown on a results page, and whether the top domain has changed. Feature Snippet is an indicator variable that takes on a value of 1 if the query is associated with a featured snippet at the top of the page, 0 otherwise. SERP Features Count is the sum of the unique number of SERP features present on the given query's results page. SERP features include carousels, top stories, reviews, videos, knowledge panels, and featured snippets.[23] In general, SERP features take up top-of-page space, crowding out advertising, thus potentially affect auctions.

[23] For a complete list of all features, see `https://developer.semrush.com/api/v3/analytics/basic-docs/#serp-features/`

Finally, Domain Change is an indicator that takes on the value of 1 when a month's top URL is different than the previous month's top URL, 0 otherwise.

Using our primary identification strategy, we test for changes in these front-end organic rank changes. We report these results in Table 7. We find that there is a roughly 11.5% increase in the likelihood a query contains a featured snippet, the number of SERP features is generally increasing by about 3.7%, and there is a 4.3% increase in the likelihood that the top domain link changes after BERT. Given that Google said that BERT would meaningfully change the result pages of about 10% of queries and increase the presence of featured snippets, these estimates seem to capture changes in organic search rankings that are consistent with Google's statement.

Table 7: Changes in organic results.

	Dependent variable:		
	Feature Snippet	log(SERP Features Count)	Domain Change
	(1)	(2)	(3)
Post × Treated	0.115^{***}	0.037^{***}	0.043^{***}
	(0.004)	(0.004)	(0.005)
Observations	104,228	104,228	104,176
R^2	0.565	0.655	0.207

Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: Estimated with competitive queries. Regressions include year-group, month, and query fixed effects. Standard errors clustered at the query level are reported in parentheses.

To test for this possibility, we estimate re-estimate Equation 3 but controlling for organic result changes. We present these results in Tables 8 and 9 for CS and CPC, respectively.

We find that the CS estimates are almost the same as those reported in Table 3. For CPC, we find that controlling for organic rank variables slightly attenuates the estimates reported in Table 4.

Table 8: Changes in $log(CS)$ by query length, controlling for organic rank changes.

	(1) Short	(2) Medium	(3) Long
Post × Treated	0.013***	0.009***	0.014***
	(0.001)	(0.001)	(0.002)
Feature Snippet	0.001	−0.001	0.005***
	(0.003)	(0.002)	(0.002)
Top Domain Change	0.001	−0.00003	0.001
	(0.001)	(0.001)	(0.001)
SERP Features Count	0.0001	0.0002	−0.002**
	(0.001)	(0.0005)	(0.001)
Observations	30,546	57,048	16,582
R^2	0.949	0.969	0.934

Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: Estimated with competitive queries. Regressions include year-group, month, and query fixed effects. Standard errors clustered at the query level are reported in parentheses.

Changes in Consumer and Advertiser Behavior Another potential explanation is that BERT affects consumer search behavior, which, in turn, affects CPC and competition. For example, consider the case in which consumers learn that results are getting better for longer/complex queries because of BERT, and this changes the types of queries consumers are searching for. Advertises, then, may start targeting and bidding for these newly popular queries, which can lead to changes in competition and CPC.

First, it is worth noting that our analysis is over a fixed set of competitive queries. Therefore, it is unlikely that our estimates directly capture this type of effect involving new search queries. However, it could be the case that the competitive queries in our dataset are indirectly affected by different search patterns through spillovers. If this was the case, however, we argue that competition for our queries should have decreased, not increased; and in response to competition changes, prices would have decreased for all queries.

Table 9: Changes in $log(CPC)$ by query length, controlling for organic rank changes.

	(1) Short	(2) Medium	(3) Long
Post × Treated	0.024^{**}	-0.018^{**}	-0.030^{**}
	(0.010)	(0.008)	(0.015)
Feature Snippet	-0.008	-0.018^{*}	-0.017
	(0.015)	(0.010)	(0.017)
Domain Change	-0.014^{**}	0.005	-0.009
	(0.006)	(0.005)	(0.010)
SERP Features Count	0.014^{***}	0.008^{**}	0.029^{***}
	(0.004)	(0.003)	(0.007)
Observations	30,546	57,048	16,582
R^2	0.696	0.736	0.690

Significance Levels: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.
Note: Estimated with competitive queries. Regressions include year-group, month, and query fixed effects. Standard errors clustered at the query level are reported in parentheses.

Second, recall that both our theoretical model and empirical analysis focus on short-term effects, which reduces the likelihood of consumers changing search behavior and advertisers responding to these changes.[24]

Third, we perform a test to partially address this concern. We collected Google Trends search interest and search volume data from SEMRush for each query in our dataset. We use these variables to account for each query's popularity.

We then re-estimate Equation 3, including these variables as control. We report these results in Tables 10 and 11. We find that controlling for Google trends and estimated search volumes does not affect the estimates.

[24]In line with the hypothesis that advertiser changes are unlikely to be happening in the short term, Alcobendas and Zeithammer (2021) find that it takes months for major advertisers to change bidding strategies following a shift from a second to a first-price auction format. In addition, from Aridor et al. (2024) also find that advertisers appear to take several months to adjust advertising budgets following data privacy regulation shifts.

Table 10: The effect of BERT on $log(CS)$ by query length controlling for search interest and search volume.

	(1) Short	(2) Medium	(3) Long
Post × Treated	0.012***	0.010***	0.013***
	(0.001)	(0.001)	(0.002)
Interest	0.0001	0.00002	0.0001
	(0.0001)	(0.00005)	(0.0001)
log(SV)	0.011***	0.003	0.003
	(0.004)	(0.003)	(0.003)
Observations	47,864	88,840	23,807
R^2	0.951	0.970	0.919

Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: Estimated with competitive queries. Regressions include year-group, month, and query fixed effects. Standard errors clustered at the query level are reported in parentheses.

6 Conclusion

Advancements in NLP research have significantly improved search engine programmatic interpretation abilities. Yet, little is known about how changes to query interpretation algorithms can affect sponsored search markets. In this paper, we develop a theoretical auction model to understand how such changes can impact the market. In our model, a seller uses an algorithm to generate signals about query types. Queries are uniquely defined by both a topic and context. Context is a distinct set of information that is fundamentally different from a query's topic and is only present in certain queries. Signals across these dimensions are used to estimate advertiser relevancy scores. Our model finds that 1) queries generally see more bidders competing when a platform improves the quality of its interpretation algorithm and 2) when improvements to contextual understanding are large, and context matters to the CTR of a query, prices may decline due to more precise relevancy scores differenti-

Table 11: The effect of BERT on $log(CPC)$ by query length controlling for search interest and search volume.

	(1) Short	(2) Medium	(3) Long
Post × Treated	0.032***	−0.017***	−0.041***
	(0.008)	(0.006)	(0.013)
Interest	−0.0002	0.0001	0.0005
	(0.0003)	(0.0003)	(0.0005)
log(SV)	0.065***	0.041***	0.102***
	(0.013)	(0.009)	(0.018)
Observations	47,864	88,840	23,807
R^2	0.686	0.728	0.670

Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: Estimated with competitive queries. Regressions include year-group, month, and query fixed effects. Standard errors clustered at the query level are reported in parentheses.

ating bidders within the auction. The former result is driven by topic signal improvements improving market allocation, while the latter is driven by contextual signal improvements helping the platform estimate more precise relevancy scores. Ultimately, the platform may be okay with lower prices as it leads to more relevant ads and higher CTR.

We then test our model predictions by studying how Google's October 2019 rollout of BERT affected sponsored search auction competition and prices. We find significant market changes: the average number of bidders uniformly increases across queries, the supply of advertising space increases, prices decline for context-rich queries and increase for context-poor queries. These findings are supported by two identification strategies and several robustness checks and are consistent with our model's predictions.

Advertisers are frequently left in the dark when search engines like Google release significant platform updates. Our paper provides a rigorous theoretical and empirical study to help practitioners and academics better understand how changes to query interpretation algorithms impact sponsored search markets. Our paper also contributes to the growing literature on the economic consequences of AI and LLMs in the search engine ecosystem. As LLMs continue to develop and increase in sophistication, our paper sheds light on the types of information these models are learning and what implications this has for markets.

References

Ada, Sıla, Nadia Abou Nabout, and Elea McDonnell Feit, "Context information can increase revenue in online display advertising auctions: Evidence from a policy change," *Journal of Marketing Research*, 2022, *59* (5), 1040–1058.

Alcobendas, Miguel and Robert Zeithammer, "Adjustment of Bidding Strategies After a Switch to First-Price Rules," *Available at SSRN 4036006*, 2021.

Amaldoss, Wilfred, Kinshuk Jerath, and Amin Sayedi, "Keyword management costs and "broad match" in sponsored search advertising," *Marketing Science*, 2016, *35* (2), 259–274.

_ , Preyas S Desai, and Woochoel Shin, "Keyword search advertising and first-page bid estimates: A strategic analysis," *Management Science*, 2015, *61* (3), 507–519.

Aridor, Guy, Yeon-Koo Che, Brett Hollenbeck, Maximilian Kaiser, and Daniel McCarthy, "Evaluating The Impact of Privacy Regulation on E-Commerce Firms: Evidence from Apples App Tracking Transparency," 2024.

Athey, Susan and Joshua S Gans, "The impact of targeting technology on advertising markets and media competition," *American Economic Review*, 2010, *100* (2), 608–613.

Bergemann, Dirk, Tibor Heumann, and Stephen Morris, "Selling Impressions: Efficiency vs. Competition," 2021.

Berman, Ron and Zsolt Katona, "The role of search engine optimization in search marketing," *Marketing Science*, 2013, *32* (4), 644–651.

Blake, Thomas, Chris Nosko, and Steven Tadelis, "Consumer heterogeneity and paid search effectiveness: A large-scale field experiment," *Econometrica*, 2015, *83* (1), 155–174.

Blei, David M, Andrew Y Ng, and Michael I Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, 2003, *3* (Jan), 993–1022.

Board, Simon, "Revealing information in auctions: the allocation effect," *Economic Theory*, 2009, *38* (1), 125–135.

Bollinger, Bryan, Eli Liebman, David Hammond, Erin Hobin, and Jocelyn Sacco, "Educational campaigns for product labels: Evidence from on-shelf nutritional labeling," *Journal of Marketing Research*, 2022, *59* (1), 153–172.

Broder, Andrei, Peter Ciccolo, Evgeniy Gabrilovich, Vanja Josifovski, Donald Metzler, Lance Riedel, and Jeffrey Yuan, "Online expansion of rare queries for sponsored search," in "Proceedings of the 18th international conference on World wide web" 2009, pp. 511–520.

Burtch, Gordon, Dokyun Lee, and Zhichen Chen, "The consequences of generative ai for ugc and online community engagement," *Available at SSRN 4521754*, 2023.

Chandra, Ambarish, "Targeted advertising: The role of subscriber characteristics in media markets," *The Journal of Industrial Economics*, 2009, *57* (1), 58–84.

Coopmans, Cas W, Karthikeya Kaushik, and Andrea E Martin, "Hierarchical structure in language and action: A formal comparison.," *Psychological Review*, 2023, *130* (4), 935.

Cowgill, Bo and Cosmina Dorobantu, "Competition and Specificity in Market Design: Evidence from Geotargeted Advertising," *Columbia Business School Research Paper*, 2020, (101).

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

Edelman, Benjamin and Zhenyu Lai, "Design of search engine services: Channel interdependence in search engine results," *Journal of Marketing Research*, 2016, *53* (6), 881–900.

_ , Michael Ostrovsky, and Michael Schwarz, "Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords," *American economic review*, 2007, *97* (1), 242–259.

Eichenbaum, Martin S, Miguel Godinho de Matos, Francisco Lima, Sergio Rebelo, and Mathias Trabandt, "Expectations, Infections, and Economic Activity," Technical Report, National Bureau of Economic Research 2020.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "Gpts are gpts: An early look at the labor market impact potential of large language models," *arXiv preprint arXiv:2303.10130*, 2023.

Ganuza, Juan-José, "Ignorance promotes competition: an auction model with endogenous private valuations," *RAND Journal of Economics*, 2004, pp. 583–598.

Ghose, Anindya and Sha Yang, "An empirical analysis of search engine advertising: Sponsored search in electronic markets," *Management science*, 2009, *55* (10), 1605–1622.

Gleason, Jeffrey, Desheng Hu, Ronald E Robertson, and Christo Wilson, "Google the gatekeeper: How search components affect clicks and attention," in "Proceedings of the International AAAI Conference on Web and Social Media," Vol. 17 2023, pp. 245–256.

Goli, Ali and Amandeep Singh, "Frontiers: Can Large Language Models Capture Human Preferences?," *Marketing Science*, 2024.

Hewitt, John and Christopher D Manning, "A structural probe for finding syntax in word representations," in "Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)" 2019, pp. 4129–4138.

Hummel, Patrick and R Preston McAfee, "When does improved targeting increase revenue?," *ACM Transactions on Economics and Computation (TEAC)*, 2016, *5* (1), 1–29.

Jäger, Gerhard and James Rogers, "Formal language theory: refining the Chomsky hierarchy," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2012, *367* (1598), 1956–1970.

Karvelis, Petros, Dimitris Gavrilis, George Georgoulas, and Chrysostomos Stylios, "Topic recommendation using Doc2Vec," in "2018 International Joint Conference on Neural Networks (IJCNN)" IEEE 2018, pp. 1–6.

Kushwaha, Amit Kumar and Arpan Kumar Kar, "MarkBot–a language model-driven chatbot for interactive marketing in post-modern world," *Information systems frontiers*, 2021, pp. 1–18.

Levin, Jonathan and Paul Milgrom, "Online advertising: Heterogeneity and conflation in market design," *American Economic Review*, 2010, *100* (2), 603–07.

Liaukonyte, Jura, Anna Tuchman, and Xinrong Zhu, "Spilling the Beans on Political Consumerism: Do Social Media Boycotts and Buycotts Translate to Real Sales Impact?," *Available at SSRN 4006546*, 2022.

Lin, Yongjie, Yi Chern Tan, and Robert Frank, "Open sesame: Getting inside BERT's linguistic knowledge," *arXiv preprint arXiv:1906.01698*, 2019.

Liu, Nelson F, Matt Gardner, Yonatan Belinkov, Matthew E Peters, and Noah A Smith, "Linguistic knowledge and transferability of contextual representations," *arXiv preprint arXiv:1903.08855*, 2019.

Lu, Yan, Michael Chau, and Patrick YK Chau, "Are sponsored links effective? Investigating the impact of trust in search engine advertising," *ACM Transactions on Management Information Systems (TMIS)*, 2017, *7* (4), 1–33.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

Milgrom, Paul R and Robert J Weber, "A theory of auctions and competitive bidding," *Econometrica: Journal of the Econometric Society*, 1982, pp. 1089–1122.

Miyagawa, Shigeru, Robert C Berwick, and Kazuo Okanoya, "The emergence of hierarchical structure in human language," *Frontiers in psychology*, 2013, *4*, 40804.

Mnih, Andriy and Geoffrey E Hinton, "A scalable hierarchical distributed language model," *Advances in neural information processing systems*, 2008, *21.*

Reisenbichler, Martin, Thomas Reutterer, David A Schweidel, and Daniel Dan, "Frontiers: Supporting content marketing with natural language generation," *Marketing Science*, 2022, *41* (3), 441–452.

Rogers, Anna, Olga Kovaleva, and Anna Rumshisky, "A primer in BERTology: What we know about how BERT works," *Transactions of the Association for Computational Linguistics*, 2021, *8*, 842–866.

Rutz, Oliver J and Randolph E Bucklin, "From generic to branded: A model of spillover in paid search advertising," *Journal of Marketing Research*, 2011, *48* (1), 87–102.

Sahni, Navdeep S and Charles Zhang, "Are consumers averse to sponsored messages? The role of search advertising in information discovery," *Quantitative Marketing and Economics*, 2024, *22* (1), 63–114.

Simonov, Andrey, Chris Nosko, and Justin M Rao, "Competition and crowd-out for brand keywords in sponsored search," *Marketing Science*, 2018, *37* (2), 200–215.

Tadelis, Steven and Florian Zettelmeyer, "Information disclosure as a matching mechanism: Theory and evidence from a field experiment," *American Economic Review*, 2015, *105* (2), 886–905.

Tenney, Ian, Dipanjan Das, and Ellie Pavlick, "BERT rediscovers the classical NLP pipeline," *arXiv preprint arXiv:1905.05950*, 2019.

_ , Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R Thomas McCoy, Najoung Kim, Benjamin Van Durme, Samuel R Bowman, Dipanjan Das et al., "What do you learn from context? probing for sentence structure in contextualized word representations," *arXiv preprint arXiv:1905.06316*, 2019.

Varian, Hal R, "Position auctions," *international Journal of industrial Organization*, 2007, *25* (6), 1163–1178.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, 2017, *30.*

Yang, Sha and Anindya Ghose, "Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence?," *Marketing science*, 2010, *29* (4), 602–623.

Zarifhonarvar, Ali, "Economics of chatgpt: A labor market view on the occupational impact of artificial intelligence," *Journal of Electronic Business & Digital Economics*, 2023.

Appendix

A Theory Model Proofs

Bidding Strategies Advertiser private valuations for a click are drawn $v_i \sim U[0,1]$. Conditional on being allocated to an auction, their objective is to maximize $E[v - \frac{\tilde{b}\tilde{r}}{r} | b^* r \geq \tilde{b}\tilde{r}]$, where b^* is the optimal bid, r is their relevancy score for the given auction, $\tilde{b}$ is the second highest bid value, and $\tilde{r}$ is the second highest relevancy score. ($\tilde{b}\tilde{r}$ is the second highest ad rank). They aim to maximize the following:

$$\int_0^{\frac{b^* r}{\tilde{r}}} (v - \frac{\tilde{Adr}}{r}) f(\tilde{b}) d\tilde{b}$$

Taking the derivative w.r.t b^* and rearranging, we arrive to $b^* = v$ $\square$

Platform Objective The platform wishes to maximize Equation 5.

$$\pi = I_{\{Click\}} \frac{\tilde{Adr}}{r_w} - I_{\{\theta_w \neq t\}} C \tag{5}$$

It will do so by selecting the optimal bidders given the information set $\hat{Q}_k$ it has available to it. We now go through the possible information sets the platform can receive and identify the optimal selection strategy within each set.

A.1 Full Information

With full information of both topic and context, the platform has two options: Allocate only topically relevant advertisers, or allocate all advertisers with non-zero relevancy scores. Let $a = 1 - B_t$ and $b = B_t$.

A.1.1 Only Topically Relevant Allocation

With only topically relevant advertisers, there are two advertisers in the auction. One advertiser has a relevancy score of 1, the other has a relevancy score of B_t.

$$(1 - B_t)\frac{B_t}{2} + B_t[\frac{1}{2}\frac{B_t}{3} + \frac{1}{2}\frac{B_t}{3}]$$

$$\frac{B_t(1 - B_t)}{2} + B_t[\frac{B_t}{3}]$$

$$= \frac{ab}{2} + \frac{b^2}{3}$$

In this scenario, the platform is allocating 2 bidders to the auction.

A.1.2 Contextually and Topically Relevant Allocation

With this scenario, the platform allocates 3 bidders to the auction. Consider the case when the $Q = (0, 0)$. The possible advertisers to allocate are $(0, 0)$, $(0, 1)$, and $(1, 0)$. (A type $(1, 1)$ advertiser has a relevancy score of 0 and will be ignored). The subsequent relevancy scores for each advertiser is 1, B_t, and $1 - B_t$. In this environment there are 5 possible cases. Define advertiser $(0, 0)$ Ad Rank as variable X, $(0, 1)$ Ad Rank as Z, and $(1, 0)$ Ad Rank as Y. $X \sim U[0, 1]$, $Z \sim U[0, b]$, and $Y \sim U[0, a]$.

With these bidders, there are five possible cases.

1. $a \leq X \leq b$, $a \leq Z \leq b$, $Y \leq a$. Occurs w.p. $\frac{(b-a)^2}{b}$
2. $X, Y, Z \leq a$. Occurs w.p. $\frac{a^2}{b}$
3. $X \geq b$. Occurs w.p. $1 - b$
4. $X, Y \leq a$, $a \leq Z \leq b$. Occurs w.p. $\frac{a(b-a)}{b}$
5. $Y, Z \leq a$, $a \leq X \leq b$. Occurs w.p. $\frac{a(b-a)}{b}$

Case 1 $a \leq X \leq b$, $a \leq Z \leq b$, $Y \leq a$. Simply a 50/50 chance either X or Z wins. If X wins, they pay Z, if Z wins, they pay X. For two uniform random variables, the expected second highest value takes on the form $d + (m-d)\frac{n-1}{n+1}$.

$$\begin{aligned}
&\frac{(b-a)^2}{b}(a + (b-a)\frac{1}{3}) \\
&= \frac{a(b-a)^2}{b} + \frac{(b-a)^3}{3b}
\end{aligned}$$

Case 2 $X, Y, Z \leq a$. Each bidder has a $\frac{1}{3}$ chance of winning within this region. If the $(1,0)$ advertiser, Y, wins, the platform pays a cost of $-C$.

$$\begin{aligned}
&\frac{a^2}{b}(\frac{1}{3}1\frac{E[max(Z,Y)]}{1} + \frac{1}{3}b\frac{E[max(X,Y)]}{b} + \frac{1}{3}(a\frac{E[max(X,Z)]}{a} - C)) \\
&= \frac{a^2}{3b}(\frac{a}{2} + \frac{a}{2} + \frac{a}{2} - C) \\
&= \frac{a^2}{3b}(\frac{3a}{2} - C)
\end{aligned}$$

Case 3 $X \geq b$. Occurs w.p. $1 - b = a$. In this case, X advertiser $(0,0)$ wins with certainty. The expected price to pay will be $\frac{max(Z,Y)}{1}$. In this scenario, Z and Y follow different distributions ($U[0,a]$ and $U[0,b]$). We need distribution of the max of the two variables to then calculate the expected value. Let $W = max(Z,Y)$

$$F(w) = Pr(Z \leq w)Pr(Y \leq w)$$

$$F(w) = \begin{cases} \frac{w^2}{ab} & w \leq a \\ \frac{w}{b} & a \leq w \leq b \\ 1 & w \geq b \end{cases}$$

$$f(w) = \begin{cases} \frac{2w}{ab} & w \leq a \\ \frac{1}{b} & a \leq w \leq b \\ 0 & w \geq b \end{cases}$$

$$E[W] = \int_0^b wf(w)dw = \int_0^a \frac{2w^2}{ab}dw + \int_a^b \frac{w}{b}dw$$
$$= \frac{2a^3}{3ab} + \frac{b}{2} - \frac{a^2}{2b}$$
$$= \frac{2a^2}{3b} + \frac{b}{2} - \frac{a^2}{2b}$$

Now the expected probability of this event is a or $1 - B_t$. In this case, X wins and has a CTR and relevancy score of 1. Final expected profit in this scenario is

$$\frac{a(3b^2 - a^2)}{6b}$$

Which simplifies to

$$\frac{ab}{2} - \frac{a^3}{6b}$$

Case 4 $X, Y \leq a$, $a \leq Z \leq b$. Occurs w.p. $\frac{a(b-a)}{b}$. In this case Z advertiser $(0, 1)$ wins. Expected profits in this scenario are

$$\frac{a(b-a)}{b}E[max(X, Y)]$$

Since X and Y are both conditionally under value a, the expected max is $\frac{2a}{3}$. This gives us a final value of

$$\frac{2a^2(b-a)}{3b}$$

Case 5 This case is identical to Case 4. Final expected profits in this scenario are

$$\frac{2a^2(b-a)}{3b}$$

Expected Profit Summing across these scenarios we reach an expected profit of

$$\frac{a^2b + b^3 - a^3}{3b} - \frac{Ca^2}{3b} + \frac{ab}{2}$$

Comparison of Profits When is it better to "misallocate" only contextually relevant bidders? When the expected profits are greater compared to only allocating relevant bidders.

$$\frac{a^2b + b^3 - a^3}{3b} - \frac{Ca^2}{3b} + \frac{ab}{2} \geq \frac{ab}{2} + \frac{b^2}{3}$$

Rearranging and simplifying, we find a simple solution that its more profitable for the platform to allocate both topically and contextually relevant bidders when $B_t \geq C$.

A.2 No Information

Now consider the case when the platform learns no information about the query. The question is whether to run an auction, and if so, how many bidders should be allocated? Note that because the platform has no information, all relevancy scores for all advertisers are equivalent and equal to $\frac{1}{2}$.

Consider the scenario in which the platform allocates all four advertisers and lets the market decide for itself.

$$E[\pi] = \frac{1}{4}(\frac{3}{5} + \frac{3B_t}{5} + \frac{3(1-B_t)}{5} - 2C)$$

$$\frac{1}{4}(\frac{6}{5} - 2C)$$

When is this greater than 0?

$$\frac{1}{4}(\frac{6}{5} - 2C) \geq 0$$
$$\frac{3}{5} \geq C$$

When $C \leq \frac{3}{5}$, the platform is willing to run an auction with all bidders. The platform has no incentive to allocate fewer advertisers as it doesn't know who is better or worse. Randomly choosing 3 advertisers from the set of 4 leads to expected profits of $\frac{5}{24} - \frac{C}{2}$ which is always less than profit with 4 bidders.

A.3 Partial Information

A.3.1 Topical Understanding, no Contextual Understanding

Consider now the case where the platform learns the query's topic, but not the context of the query. Here $\hat{t} = t$ and $\hat{q} = \frac{1}{2}$. Consider a query $Q = (0, 0)$ and all four advertiser $X = (0, 0)$, $Y = (0, 1)$, $Z = (1, 0)$, and $U = (1, 1)$. The relevancy scores for X and Y advertisers are $\frac{1+B_t}{2}$ and $\frac{1-B_t}{2}$ for Z and U advertisers. Topically relevant advertisers are X and Y.

Allocate Only Topically Relevant If the platform allocates only topically relevant advertisers X and Y, it creates a simple second-price auction outcome. (Relevancy scores cancel out, so the model simplifies to a simple SPA). There is a $\frac{1}{2}$ chance a given advertiser wins. If X wins, CTR is 1. If Y wins, CTR is B_t. Expected second highest value is $\frac{1}{3}$.

$$E[\pi] = \frac{1}{6} + \frac{B_t}{6}$$
$$= \frac{1 + B_t}{6}$$

Allocate All Advertisers Now, the platform could allocate all advertisers with non-zero relevancy scores. (If $B_t = 1$, Z and U advertisers are essentially removed). Define $a = \frac{1-B_t}{2}$

and $b = \frac{1+B_t}{2}$. There are four possible cases that could occur with all four advertisers allocated to the auction.

1. $X, Y, Z, U \leq a$. Occurs w.p. $\frac{a^2}{b^2}$
2. $X, Y \in (a, b]$ and $Z, U \leq a$. Occurs w.p. $\frac{(b-a)^2}{b^2}$
3. $X \geq a$ and $Y, Z, U \leq a$. Occurs w.p. $\frac{a(b-a)}{b^2}$
4. $Y \geq a$ and $X, Z, U \leq a$. Occurs w.p. $\frac{a(b-a)}{b^2}$

Case 1: The first case considers when all four bidder ad ranks fall under a. In this scenario, each bidder has an equal chance of winning. Define W as the expected second highest ad rank.

$$\frac{a^2}{b^2}[\frac{1}{4}X + \frac{1}{4}Y + \frac{1}{4}Z + \frac{1}{4}U]$$

$$\frac{a^2}{b^2}[\frac{1}{4}\frac{W}{b} + \frac{1}{4}\frac{W}{b}B_t + \frac{1}{4}((1 - B_t)\frac{W}{a} - C) - \frac{1}{4}C]$$

$$\frac{a^2}{b^2}[W - \frac{C}{2}]$$

Now W is the expected second highest value among 4 draws from $U[0, a]$. This translates to $\frac{3a}{5}$.

$$\frac{a^2}{b^2}[\frac{3a}{5} - \frac{C}{2}]$$

Case 2: In this case, X and Y are in the range of a and b. This occurs w.p. $\frac{(b-a)^2}{b^2}$. Define W as now the second highest value among X and Y, conditional on X and Y being greater than a. Both X and Y have a relevancy score of b.

$$\frac{(b-a)^2}{b^2}[\frac{1}{2}X + \frac{1}{2}Y]$$

$$\frac{(b-a)^2}{b^2}[\frac{W}{2b} + \frac{W}{2b}B_t]$$

$$\frac{(b-a)^2}{b^2}[\frac{W}{2b}(1 + B_t)]$$

Since $2b = 1 + B_t$

$$\frac{(b-a)^2}{b^2}[W]$$

Now W is the second highest value among two draws from $U[a, b]$. This translates to $\frac{1}{3}(b - a) + a$.

$$\frac{(b-a)^2}{b^2}(\frac{b-a}{3} + a)$$

Case 3: This case occurs when $X \geq a$ and all other ad ranks are less than or equal to a. This occurs w.p. $\frac{a(b-a)}{b^2}$. In this case, X wins and pays the max among Y, Z, and U scaled by its relevancy score. Define W as the expected max among Y, Z, U. It is equal to $\frac{3a}{4}$

$$\frac{a(b-a)}{b^2}[\frac{W}{b}]$$

$$\frac{a(b-a)}{b^2}[\frac{3a}{4b}]$$

$$\frac{3a^2(b-a)}{4b^3}$$

Case 4: Case 4 is similar to case 3, except expected CTR for Y is B_t instead of 1 like it is for X. Therefore, we have

$$\frac{a(b-a)}{b^2}[\frac{3a}{4b}B_t]$$

Expected Profits This covers all the possible cases. Expected profit is now the summation of these cases. First, consider that Case 3 and 4 joined are:

$$(1+B_t)\frac{3a^2(b-a)}{4b^3}$$

Since $2b = (1+B_t)$, we can simplify this to

$$\frac{3a^2(b-a)}{2b^2}$$

Summing all cases now:

$$E[\pi] = \frac{a^2}{b^2}[\frac{3a}{5} - \frac{C}{2}] + \frac{(b-a)^2}{b^2}[\frac{b-a}{3} + a] + \frac{3a^2(b-a)}{2b^2}$$

What we want to know is when is this more profitable in expectation than allocating just the topically relevant advertisers.

$$\frac{a^2}{b^2}[\frac{3a}{5} - \frac{C}{2}] + \frac{(b-a)^2}{b^2}[\frac{b-a}{3} + a] + \frac{3a^2(b-a)}{2b^2} \geq \frac{b}{3}$$

Simplifying, we get:

$$b - \frac{7}{15}a \geq C$$

Since $b = \frac{1+B_t}{2}$ and $a = \frac{1-B_t}{2}$, we can further simplify

$$\frac{4+11B_t}{15} \geq C$$

When $C \leq \frac{4+11B_t}{15}$, it is more profitable for the company to allocate all advertisers to the auction than to run the auction with only topically relevant advertisers.

Allocate Topically Relevant and a Topically Irrelevant Now, there could be cases where the platform only wants to allocate topically relevant and one topically irrelevant advertiser. The platform may benefit from this if it can weakly boost the final price paid. Here, they can allocate X and Y and either Z or U. If they allocate Z or U, there's a chance they're contextually relevant. Again define $a = \frac{1-B_t}{2}$ and $b = \frac{1+B_t}{2}$.

1. $X, Y \geq a$, $Z/U \leq a$. Occurs w.p. $\frac{(b-a)^2}{b^2}$
2. $X \geq a$, $Y, Z/U \leq a$. Occurs w.p. $\frac{a(b-a)}{b^2}$
3. $Y \geq a$, $X, Z/U \leq a$. Occurs w.p. $\frac{a(b-a)}{b^2}$
4. $X, Y, Z/Y \leq a$. Occurs w.p. $\frac{a^2}{b^2}$

Case 1: In this scenario, the effect of Z/U on price is irrelevant. Expected profits are the same as Case 2 in the previous option when all advertisers are allocated.

$$\frac{(b-a)^2}{b^2}(\frac{b-a}{3} + a)$$

Case 2: Here, X wins and pays the second highest price, or the max between Y and the other advertiser Z/U. Define W as the max of the two variables Y and Z/U.

$$\frac{a(b-a)}{b^2}[1\frac{W}{b}]$$

$$\frac{a(b-a)}{b^2}[\frac{2a}{3b}]$$

$$\frac{2a^2(b-a)}{3b^3}$$

Case 3: This is the same as Case 2, except expected CTR is B_t instead of 1.

$$\frac{2a^2(b-a)}{3b^3}B_t$$

Case 4: The final case considers when X, Y, and Z/U are all less than or equal to a. Define W as the expected second highest draw among the three variables between 0 and a.

$$\frac{a^2}{b^2}(\frac{1}{3}X + \frac{1}{3}Y + \frac{1}{3}Z/U)$$

$$\frac{a^2}{b^2}(\frac{1}{3}\frac{W}{b} + \frac{1}{3}\frac{W}{b}B_t + \frac{1}{3}[\frac{1}{2}(\frac{W}{a}(1 - B_t) - C) - \frac{1}{2}C])$$

$$\frac{a^2}{b^2}(\frac{W}{3b}(1 + B_t) + \frac{1}{6}(\frac{W}{a}(1 - B_t) - 2C))$$

Since $2a = (1 - B_t)$ and $2b = (1 + B_t)$, we can further simplify.

$$\frac{a^2}{b^2}(\frac{2W}{3} + \frac{1}{6}(2W - 2C))$$

$$\frac{a^2}{b^2}(W - \frac{C}{3})$$

Now W is the second highest draw of 3 draws from the distribution $U[0, a]$. This translates to $\frac{a}{2}$. This gives us:

$$\frac{a^2}{b^2}(\frac{a}{2} - \frac{C}{3})$$

Summing Up Cases

$$\frac{(b-a)^2}{b^2}(\frac{b-a}{3} + a) + \frac{2a^2(b-a)}{3b^3} + \frac{2a^2(b-a)}{3b^3}B_t + \frac{a^2}{b^2}(\frac{a}{2} - \frac{C}{3})$$

$$\frac{(b-a)^2}{b^2}(\frac{b-a}{3} + a) + \frac{4a^2(b-a)}{3b^2} + \frac{a^2}{b^2}(\frac{a}{2} - \frac{C}{3})$$

Comparison Across Selection Choices We want to know when this decision is more profitable than allocating only topically relevant or allocating all advertisers. First, consider case for when expected profit is greater than allocating only topically relevant advertisers.

$$\frac{(b-a)^2}{b^2}(\frac{b-a}{3}+a)+\frac{4a^2(b-a)}{3b^2}+\frac{a^2}{b^2}(\frac{a}{2}-\frac{C}{3})\geq\frac{b}{3}$$

Simplifying, we get the following:

$$\frac{1+3B_t}{4}\geq C$$

So, when $C\leq\frac{1+3B_t}{4}$, its more profitable to allocate topically relevant and one topically irrelevant advertiser to the auction compared to allocating only topically relevant.

Now, we also need to compare to allocating all advertisers.

$$\frac{(b-a)^2}{b^2}(\frac{b-a}{3}+a)+\frac{4a^2(b-a)}{3b^2}+\frac{a^2}{b^2}(\frac{a}{2}-\frac{C}{3})\geq\frac{a^2}{2b^2}[\frac{6a}{5}-C]+\frac{(b-a)^2}{b^2}[\frac{b-a}{3}+a]+\frac{3a^2(b-a)}{2b^2}$$

Simplifying this down, we get

$$\frac{3+7B_t}{10}\leq C$$

So when $\frac{3+7B_t}{10}\leq C$, its better to allocate 3 advertisers instead of all 4.

Allocating Two Topically Irrelevant and One Topically Relevant Advertiser The platform could allocate two topically irrelevant advertisers and one topically relevant advertisers. This would lead to an expected profit of:

$$E[\pi]=\frac{(b-a)}{b}\frac{2a}{3}+\frac{a}{b}[\frac{a}{2}-\frac{2C}{3}]$$

When simplifying, this translates to

$$\frac{1-B_t}{1+B_t}(\frac{2}{3}B_t + \frac{1-B_t}{4} - \frac{2}{3}C)$$

Compare it to when we allocate 3 bidders where 2 are topically relevant and 1 is topically irrelevant.

$$\frac{(b-a)}{b}\frac{2a}{3} + \frac{a}{b}[\frac{a}{2} - \frac{2C}{3}] \geq \frac{(b-a)^2}{b^2}(\frac{b-a}{3} + a) + \frac{4a^2(b-a)}{3b^2} + \frac{a^2}{b^2}(\frac{a}{2} - \frac{C}{3})$$

$$0 \geq 12a^2b - 4ab^2 - 5a^3 - 3ab + 2C(2ab - a^2) + 2(b-a)^3 + 6a(b-a)^2$$

This eventually simplifies to

$$(1 - 3B_t^2 + 2B_t)C \leq \frac{B_t + B_t^2 - 5B_t^3 - 5}{4}$$

Note that between 0 and 1, $\frac{B_t+B_t^2-5B_t^3-5}{4}$ is always negative. Since $C \geq 0$ and $1 - 3B_t^2 + 2B_t \geq 0$, the left hand side is always positive, so this inequality never holds. Since allocating two topically relevant and one topically irrelevant is already trumped by alternative selections, the choice to select two topically irrelevant and one topically relevant is not relevant.

What Selection is Optimal Under these conditions, we now have 3 cutoff values: when $\frac{3+7B_t}{10} \leq C$ 3 bidders is better than 4, when $\frac{1+3B_t}{4} \geq C$ 3 is better than 2, and when $\frac{4+11B_t}{15} \geq C$ 4 is better than 2.

$$\frac{1+3B_t}{4} \leq \frac{4+11B_t}{15} \leq \frac{3+7B_t}{10}, \frac{1}{2} \leq B_t \leq 1$$

When $\frac{3+7B_t}{10} \leq C$, its better to allocate 2 bidders over 3 and 4 bidders. When $\frac{4+11B_t}{15} \leq C \leq \frac{3+7B_t}{10}$ its optimally to still only allocate 2 bidders. When $\frac{1+3B_t}{4} \leq C \leq \frac{4+11B_t}{15}$, it's optimal to allocate all 4 bidders. And finally, when $C \leq \frac{1+3B_t}{4}$ its optimal to also allocate all 4 bidders.

1. $\frac{3+7B_t}{10} \leq C$ Two bidders.

2. $\frac{4+11B_t}{15} \leq C \leq \frac{3+7B_t}{10}$ Two bidders.

3. $\frac{1+3B_t}{4} \leq C \leq \frac{4+11B_t}{15}$ Four bidders.

4. $C \leq \frac{1+3B_t}{4}$ Four bidders.

So the primary cutoff value is $\frac{4+11B_t}{15}$. If $C \geq \frac{4+11B_t}{15}$, the platform allocates only the topically relevant bidders, otherwise the platform allocates all bidders. $\square$

A.3.2 Contextual Understanding, no Topical Understanding

The final partial information case to consider is when the platform learns the context but not the topic of the query. Consider a query $Q = (0, 0)$ and advertisers X, Y, Z, and U where $X, Y \sim U[0, b]$ and $Z, U \sim U[0, a]$. Advertisers X and Y have relevancy scores equal to $1 - \frac{B_t}{2}$ while Z and U advertisers have relevancy scores of $\frac{B_t}{2}$.

Allocate Only Contextually Relevant Consider the case when the platform only allocates known contextually relevant advertisers (X and Y). Both advertisers have the same relevancy score. Let W equal the second highest value from the two random variables X and Y.

$$E[\pi] = [\frac{1}{2}X + \frac{1}{2}Y]$$

$$\frac{1}{2}\frac{W}{b} + \frac{1}{2}((1 - B_t)\frac{W}{b} - C)$$
$$\frac{W}{2b}(2 - B_t) - \frac{C}{2}$$

Note that $2b = 2 - B_t$

$$W - \frac{C}{2}$$

Now, $W = \frac{b}{3}$, where $b = 1 - \frac{B_t}{2}$. We therefore have:

$$E[\pi] = \frac{b}{3} - \frac{C}{2}$$

This is the expected profit allocating only the advertisers with known context. If $C \geq \frac{2-B_t}{3}$, it is not profitable in expectation for the firm to run an auction with only contextually relevant bidders.

Allocate All Bidders An alternative option for the platform is to allocate all bidders. Let $a = \frac{B_t}{2}$ and $b = 1 - \frac{B_t}{2}$. This creates four cases.

1. $X, Y \in (a, b]$ and $Z, U \leq a$. Occurs with probability $\frac{(b-a)^2}{b^2}$
2. $X \geq a$ and $Y, Z, U \leq a$. Occurs w.p. $\frac{a(b-a)}{b^2}$
3. $Y \geq a$ and $X, Z, U \leq a$. Occurs w.p. $\frac{a(b-a)}{b^2}$
4. $X, Y, Z, U \leq a$. Occurs w.p. $\frac{a^2}{b^2}$

We now consider these cases.

Case 1: $X, Y \in (a, b]$. This occurs w.p. $\frac{(b-a)^2}{b^2}$. The expected profit is the same structure as when the platform allocates only contextually relevant bidders, except now the lower bound of values is a and not 0. Define W as the expected second highest Ad Rank among X and Y given they're greater than or equal to a.

$$\frac{(b-a)^2}{b^2}[\frac{1}{2}X + \frac{1}{2}Y]$$

$$\frac{(b-a)^2}{b^2}[\frac{1}{2}\frac{W}{b} + \frac{1}{2}((1 - B_t)\frac{W}{b} - C)]$$

$$\frac{(b-a)^2}{b^2}[W-\frac{C}{2}]$$

Note that $W=\frac{1}{3}(b-a)+a$. We find that expected profit is:

$$\frac{(b-a)^2}{b^2}[\frac{(b-a)}{3}+a-\frac{C}{2}]$$

Case 2: This occurs with $X \geq a$ and $Y \leq a$. This occurs w.p. $\frac{a(b-a)}{b^2}$. Define W as the expected max value among the three variables Y, Z, and U, given they're all less than or equal to a. Expected profit is:

$$\frac{a(b-a)}{b^2}[\frac{1}{2}\frac{W}{b}+\frac{1}{2}((1-B_t)\frac{W}{b}-C)]$$
$$\frac{a(b-a)}{b^2}[W-\frac{C}{2}]$$

Here, $W=\frac{3a}{4}$. Final expected profit in this case is

$$\frac{a(b-a)}{b^2}[\frac{3a}{4}-\frac{C}{2}]$$

Case 3: Case 3 is symmetrical to case 2, so final expected profit in this scenario is:

$$\frac{a(b-a)}{b^2}[\frac{3a}{4}-\frac{C}{2}]$$

Case 4: The final case is when all four ad ranks are below a. This occurs w.p. $\frac{a^2}{b^2}$. Each bidder has an equal likelihood of winning.

$$\frac{a^2}{b^2}[\frac{1}{4}X+\frac{1}{4}Y+\frac{1}{4}Z+\frac{1}{4}U]$$
$$\frac{a^2}{b^2}[\frac{W}{b}+(\frac{W}{b}(1-B_t)-C)+\frac{W}{a}B_t-C]$$

Here, W is defined as the expected second highest ad rank among the four bidders, conditional on their values being less than or equal to a. This again simplifies down to $W - \frac{C}{2}$

$$\frac{a^2}{b^2}[W - \frac{C}{2}]$$

Where $W = \frac{3}{5}a$

$$\frac{a^2}{b^2}[\frac{3a}{5} - \frac{C}{2}]$$

Total Expected Profit Total profit is going to be the summation of each of these events.

$$\frac{(b-a)^2}{b^2}[\frac{(b-a)}{3} + a - \frac{C}{2}] + \frac{a(b-a)}{b^2}[\frac{3a}{4} - \frac{C}{2}] + \frac{a(b-a)}{b^2}[\frac{3a}{4} - \frac{C}{2}] + \frac{a^2}{b^2}[\frac{3a}{5} - \frac{C}{2}]$$

The question is then, when is this more profitable than allocating only contextually relevant bidders or compared to not running an auction.

$$\frac{(b-a)^2}{b^2}[\frac{(b-a)}{3} + a - \frac{C}{2}] + \frac{a(b-a)}{b^2}[\frac{3a}{4} - \frac{C}{2}] + \frac{a(b-a)}{b^2}[\frac{3a}{4} - \frac{C}{2}] + \frac{a^2}{b^2}[\frac{3a}{5} - \frac{C}{2}] \geq \frac{b}{3} - \frac{C}{2}$$

Notice that $\frac{C}{2}$ is common among all outcomes and so the left and right side each have a $\frac{C}{2}$, so we can cancel it out.

$$\frac{(b-a)^2}{b^2}[\frac{(b-a)}{3} + a] + \frac{a(b-a)}{b^2}[\frac{3a}{4}] + \frac{a(b-a)}{b^2}[\frac{3a}{4}] + \frac{a^2}{b^2}[\frac{3a}{5}] \geq \frac{b}{3}$$

$$\frac{(b-a)^2}{b^2}[\frac{(b-a)}{3} + a] + \frac{a(b-a)}{b^2}[\frac{3a}{2}] + \frac{a^2}{b^2}[\frac{3a}{5}] \geq \frac{b}{3}$$

$$\frac{(1+a)(b-a)^2}{3b^2} + \frac{3a^2(b-a)}{2b^2} + \frac{3a^3}{5b^2} \geq \frac{b}{3}$$

Simplifying this inequality, we find that

$$B_t \leq \frac{15}{11}$$

In other words, it is more profitable to allocate all advertisers that to only allocate contextually relevant advertisers when $B_t \leq \frac{15}{11}$. Since $B_t \leq 1$, this always holds.

Comparing to no Auction

$$\frac{(b-a)^2}{b^2}(\frac{1+a}{3}) + \frac{a(b-a)}{b^2}\frac{3a}{2} + \frac{a^2}{b^2}\frac{3a}{5} \geq \frac{C}{2}$$

We can simplify this down to

$$C \leq \frac{10 - 30a + 45a^2 - 32a^3}{15(1-a)^2}$$

Which can be further simplified to

$$C \leq \frac{4}{(2-B_t)^2}[\frac{2}{3} - B_t + B_t^2 - \frac{4}{15}B_t^3]$$

When C is less than or equal to $\frac{4}{(2-B_t)^2}[\frac{2}{3} - B_t + B_t^2 - \frac{4}{15}B_t^3]$, it is more profitable for the company to run an auction with all bidders than to not run any auction.

Allocate Contextually Relevant and one Contextually Irrelevant Bidder Let $a = \frac{B_t}{2}$ and $b = 1 - \frac{B_t}{2}$. This creates four cases.

1. $X, Y \in (a, b]$ and $Z/U \leq a$. Occurs with probability $\frac{(b-a)^2}{b^2}$

2. $X \geq a$ and $Y, Z/U \leq a$. Occurs w.p. $\frac{a(b-a)}{b^2}$

3. $Y \geq a$ and $X, Z/U \leq a$. Occurs w.p. $\frac{a(b-a)}{b^2}$

4. $X, Y, Z/U \leq a$. Occurs w.p. $\frac{a^2}{b^2}$

Case 1:

$$E[\pi] = \frac{(b-a)^2}{b^2}(\frac{1+a}{3} - \frac{C}{2})$$

Cases 2 and 3:

$$E[\pi] = \frac{a(b-a)}{b^2}[\frac{2a}{3} - \frac{C}{2}]$$

Case 4:

$$E[\pi] = \frac{a^2}{b^2}[\frac{a}{2} - \frac{C}{2}]$$

Summing Up Cases Summing up the cases we have

$$E[\pi] = \frac{(b-a)^2}{b^2}(\frac{1+a}{3}) + \frac{a(b-a)}{b^2}(\frac{4a}{3}) + \frac{a^2}{b^2}(\frac{a}{2}) - \frac{C}{2}$$

Now we want to consider when this expected profit is greater than allocating only contextually relevant bidders or allocating all bidders. Since allocating all bidders dominates allocating only contextually relevant bidders, we will compare to allocating all bidders

Comparing to Allocating All Bidders

$$\frac{(b-a)^2}{b^2}(\frac{1+a}{3}) + \frac{a(b-a)}{b^2}(\frac{4a}{3}) + \frac{a^2}{b^2}(\frac{a}{2}) - \frac{C}{2} \geq \frac{(1+a)(b-a)^2}{3b^2} + \frac{3a^2(b-a)}{2b^2} + \frac{3a^3}{5b^2} - \frac{C}{2}$$

Here, both $\frac{C}{2}$ and $\frac{(b-a)^2}{b^2}(\frac{1+a}{3})$ cancel out. This leaves us with:

$$\frac{4a^2(b-a)}{3b^2} + \frac{a^3}{2b^2} \geq \frac{3a^2(b-a)}{2b^2} + \frac{3a^3}{5b^2}$$

$$-\frac{a^2(b-a)}{6} \geq \frac{a^3}{10}$$

$$10b \leq 4a$$

$$10(1-a) \leq 4a$$

$$\frac{5}{7} \leq a$$

$$\frac{10}{7} \leq B_t$$

Therefore, When $B_t \geq \frac{10}{7}$, the platform is better off running the auction with 3 bidders and not all 4. However, $B_t \leq 1$, so this cannot occur.

Allocate 2 Non Contextually relevant and 1 Contextually Relevant The platform could consider allocating 2 non contextually relevant advertisers and only 1 contextually relevant advertiser to the auction. Expected profit in this case is:

$$E[\pi] = \frac{2a(b-a)}{3b} + \frac{a^2}{2b} - \frac{C}{2}$$

Comparing to allocating all four bidders:

$$\frac{2a(b-a)}{3b} + \frac{a^2}{2b} - \frac{C}{2} \geq \frac{(1+a)(b-a)^2}{3b^2} + \frac{3a^2(b-a)}{2b^2} + \frac{3a^3}{5b^2} - \frac{C}{2}$$

It is only better to allocate this combination of three bidders compared to all four when

$$0 \geq 10 - 25B_t + \frac{45}{2}B_t^2 - \frac{57}{8}B_t^3$$

However, this inequality does not hold for all $\frac{1}{2} \leq B_t \leq 1$. Therefore, it is not a considered platform option.

A.4 Region Analysis

We now have our selection process criteria. Under full information, when $C \leq B_t$, it is better to allocate all topically and contextually relevant bidders. When $C \geq B_t$, the platform only wants to allocate topically relevant bidders. Under no information, the platform allocates all bidders if $C \leq \frac{3}{5}$, otherwise none are allocated. When the platform knows the topic dimension, but not the context dimension, the platform allocates only topically relevant bidders when $C \geq \frac{4+11B_t}{15}$. Otherwise, it allocates all bidders. When context is understood, the platform allocates all bidders only when $C \leq \frac{4}{(2-B_t)^2}[\frac{2}{3} - B_t + B_t^2 - \frac{4}{15}B_t^3]$, otherwise no bidders are allocated. These inequalities create six unique regions that depend on C and B_t. In Figure 5 we plot the regions created by these thresholds.

Recall, that w.p. γ_t the platform learns the topic dimension of the query and w.p. γ_q, the platform learns the context dimension of the query. We now describe expected number of bidders and expected CPC based on region.

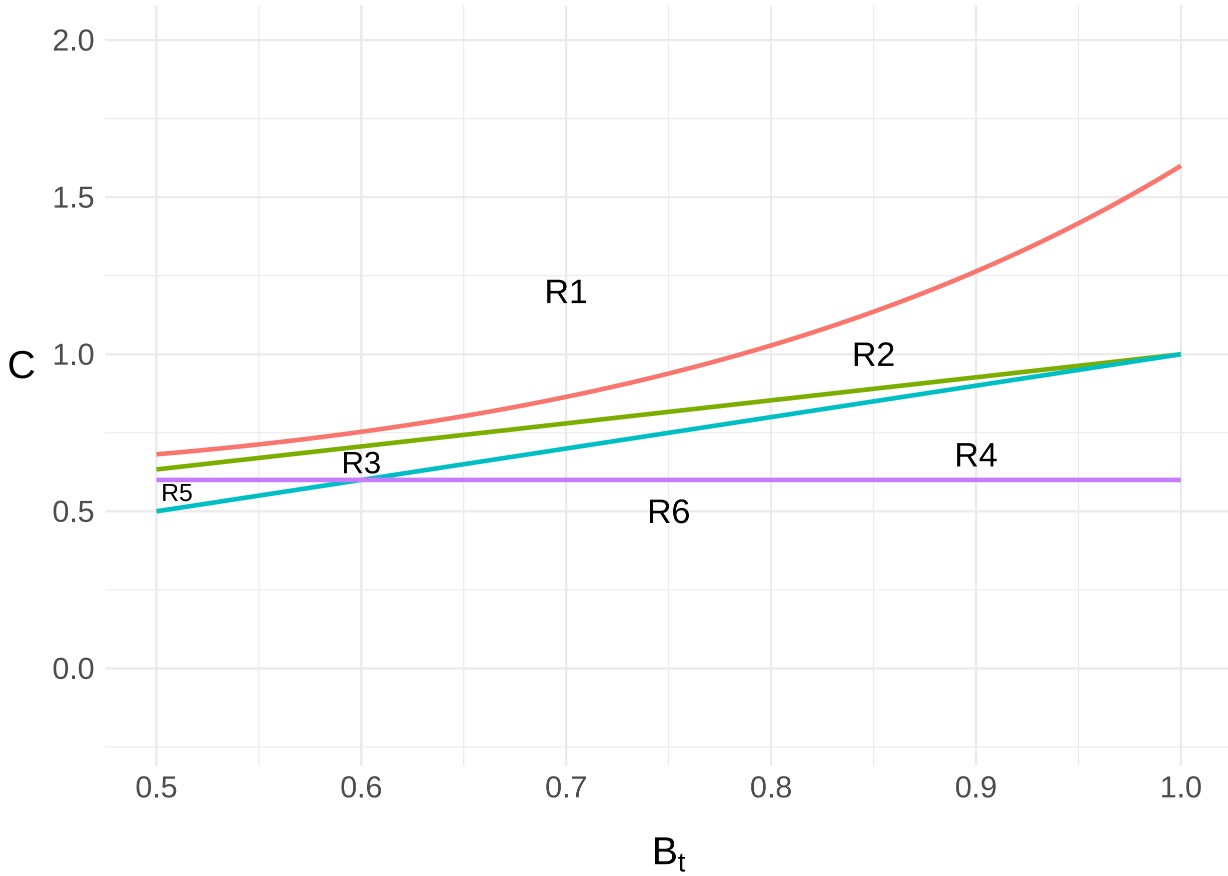
1. **Region 1:** $C \geq \frac{4}{(2-B_t)^2}[\frac{2}{3} - B_t + B_t^2 - \frac{4}{15}B_t^3]$. The platform only allocates 2 bidders when topics are known. This leads to expected number of bidders of $2\gamma_t$.

2. **Region 2:** $C \leq \frac{4}{(2-B_t)^2}[\frac{2}{3} - B_t + B_t^2 - \frac{4}{15}B_t^3]$ and $C \geq \frac{4+11B_t}{15}$. Here the platform allocates only topically relevant bidders when the topic dimension is understood and all four bidders when only the context dimension is understood. This leads to an expected number of bidders of $2\gamma_t + 4\gamma_q - 4\gamma_t\gamma_q$.

3. **Region 3:** $C \leq \frac{4+11B_t}{15}$ and $C \geq B_t$. In this region, the platform allocates 2 bidders with full information, and all four bidders when either the topic or context dimension is known. This leads to an expected number of bidders of $4(\gamma_t + \gamma_q) - 6\gamma_t\gamma_q$.

4. **Region 4:** $C \leq B_t$ and $C \geq \frac{3}{5}$. Here, the platform is willing to allocate all three bidders with non-zero relevancy scores with full information and allocate all bidders when either topic or context is known. This translates to an expected number of bidders of $4(\gamma_t + \gamma_q) - 5\gamma_t\gamma_q$.

Figure 5: Regions that determine how auction evolves with changes to a. C is the cost the platform incurs for showing a topically irrelevant ad and B_t is the importance of topic alignment for a query's CTR.



5. **Region 5:** $C \leq \frac{3}{5}$ and $C \geq B_t$. Here, the platform only wants to allocate topically relevant bidders with full information, but is willing to allocate all bidders in all other cases. This translates to an expected number of bidders of $4 - 2\gamma_t\gamma_q$.

6. **Region 6:** $C \leq \frac{3}{5}$ and $C \leq B_t$. Here, the platform allocates every bidder with non-zero relevancy scores. With full information, we assume the platform allocates all 3 bidders with non-zero relevancy scores. This translates to an expected number of bidders $4 - \gamma_t\gamma_q$.

Note that Google states it does not run auctions when it isn't confident it understands the query. Therefore, we can assume that the cost $C \geq \frac{3}{5}$ and the platform is unwilling to run auctions when there is no signal about the query's type. We will focus on regions 1 through 4 for analysis and exclude 5 and 6 for this reason.

We can notice several things from the plot and the results. First, for queries where topic alignment is the priority (i.e., short queries where B_t is large), Regions 2 and 4 dominate most of the possible spaces. For queries where context matters more, Region 1 takes up more and more of the potential space. If C is greater than 1, then all queries with B_t roughly greater than 0.7852 will be in Region 2, with all other queries in Region 1. As C lowers, we start to enter other regions. Consider when $C = 0.8$. All queries with $B_t \leq 0.647$ will live in Region 1, all queries with $0.6471 \leq B_t \leq \frac{8}{11}$ will be in Region 2, all queries where $\frac{8}{11} \leq B_t \leq 0.8$ will be in Region 3, and all queries with $B_t \geq 0.8$ will live in Region 4.

These regions define the auction environment in which a given query lives in and determines how the market will change with the introduction of a new algorithm. Note that CPC is calculated as $Pr(Click)\frac{\tilde{Adr}}{r_w}$ where $\tilde{Adr}$ is the second highest ad rank and r_w is the winning relevancy score. If no click is received, CPC is 0, if a click is received, its the second highest ad rank scaled by the winning ad rank score. We now calculate expected CPC and number of bidders for each region.

A.4.1 Region 1

Number of Bidders In Region 1, the platform is conservative and only allocates topically relevant bidders. Here the expected number of bidders is $2\gamma_t$ and so the number of bidders is increasing with a since $\frac{\partial \gamma_t(a)}{\partial a} > 0$.

Cost-Per-Click (CPC) In Region 1, the platform only allocates topically relevant bidders. Expected CPC is then

$$\gamma_q \gamma_t (\frac{(1 - B_t)B_t}{2} + \frac{B^2}{3}) + \gamma_t (1 - \gamma_q) \frac{1 + B_t}{6}$$

We can simplify to

$$\frac{\gamma_t}{6}[1 + B_t - \gamma_q(B_t - 1)^2]$$

Taking the derivative w.r.t. to a and setting less than 0, we find that CPC declines when

$$\frac{1 + B_t}{6(B_t - 1)^2} < \frac{\gamma_q^{'}}{\gamma_t^{'}}\gamma_t + \gamma_q$$

Notice that as B_t closes in on 1, the left-hand value approaches infinity, meaning it is unlikely that prices can decline. This matches to our intuition that short queries, which lack context and therefore have a higher B_t value are unlikely to see prices decline because context does not exist to further differentiate bidders within auctions. But, when B_t is moderate and gains to contextual understanding are large relative to topical gains, there is an opportunity for prices to decline. The intuition for the result is that when the platform's interpretation algorithm substantially improves context signals it leads to higher frequency of full information auction events and improved bidder differentiation. This can lead to lower prices but more relevant advertisements and higher expected profit. (Lower prices but higher CTR).

A.4.2 Region 2

Number of Bidders In Region 2, the platform allows for some misallocation and we observe the number of bidders equal to $2\gamma_t + 4\gamma_q - 4\gamma_t\gamma_q$. Taking the derivative w.r.t. γ_q equates to $4(1 - \gamma_t)$, which is always non-negative. Taking the derivative w.r.t. γ_t equates to $2 - 4\gamma_q$, which is positive when $\gamma_q < \frac{1}{2}$, 0 when $\gamma_q = \frac{1}{2}$, and negative when $\gamma_q > \frac{1}{2}$.

Taking the derivative of this w.r.t a, we find that the number of bidders will decline when:

$$(1 - \gamma_t)\frac{\gamma_q^{'}}{\gamma_t^{'}} < \gamma_q - \frac{1}{2}$$

When the inequality flips, the number of bidders is increasing and when there is equality the number of bidders stays the same with a. Note that $\frac{\gamma_q^{'}}{\gamma_t^{'}}$ is always positive and that $\gamma_q - \frac{1}{2}$

only begins to be positive when $\gamma_q > \frac{1}{2}$. Therefore, when the context signal is weak ($\gamma_q < \frac{1}{2}$), it's not possible to see a decline in the number of bidders.

When $\gamma_q > \frac{1}{2}$, it is possible to see a decline in the number of bidders. Assuming the improvements to contextual signals are small (small $\gamma_q^{'}$), a decline could occur when the topic signal is also strong (large γ_t) or improvements to topic signals are large (large $\gamma_t^{'}$).

When the platform learns the topic dimension, the market is small (two bidders). So, as this signal improves (larger γ_t), the left-hand side term shrinks due to $(1 - \gamma_t)$. Whether this leads to a decline in the average number of bidders depends on the strength of the context signal. If the context signal is large and getting all advertisers through the door in certain situations, then we can potentially see the average number of bidders decline. Intuitively, this is capturing the market segmentation effects we might imagine could occur. The platform would be, on average, removing some irrelevant advertisers and thinning the markets to more relevant advertisers.

Cost-Per-Click (CPC) Region 2 considers when the platform allocates two topically relevant when the topic is known and all four bidders when the context is known but the topic is not known. Define $J = \frac{40-60B_t+45B_t^2-16B_t^3}{30(2-B_t)^2}$. Expected CPC in Region 2 for a query is defined as

$$\frac{\gamma_t}{6}[1 + B_t - \gamma_q(B_t - 1)^2] + \gamma_q(1 - \gamma_t)A$$

To simplify notation, let $D = \frac{1+B_t}{6}$ and let $M = J + \frac{(B_t-1)^2}{6}$. Average CPC in Region 2 becomes

$$\gamma_t D + J\gamma_q - \gamma_t\gamma_q M$$

Taking the derivative w.r.t. a and finding when its less than 0, we get

$$\frac{\gamma_q^{'}}{\gamma_t^{'}}(J - M\gamma_t) < M\gamma_q - D$$

We now are interested in which cases cause this inequality to hold or not. When it does not hold, CPC can either increase or stay the same.

1. $\gamma_q < \frac{D}{M}$ and $\gamma_t \leq \frac{J}{M}$. Inequality cannot hold. CPC must increase.
2. $\gamma_q < \frac{D}{M}$ and $\gamma_t > \frac{J}{M}$. Inequality can hold. CPC may decline if $\frac{\gamma_q^{'}}{\gamma_t^{'}}$ is large.
3. $\gamma_q \geq \frac{D}{M}$ and $\gamma_t < \frac{J}{M}$. Inequality can hold. CPC may decline if $\frac{\gamma_q^{'}}{\gamma_t^{'}}$ is small.
4. $\gamma_q > \frac{D}{M}$ and $\gamma_t = \frac{J}{M}$. Inequality always holds and CPC must decline. When $\gamma_q = \frac{D}{M}$ and $\gamma_t = \frac{J}{M}$, the sides equal and implies that there is no change to CPC.
5. $\gamma_q \geq \frac{D}{M}$ and $\gamma_t > \frac{J}{M}$. Inequality always holds. CPC drops with certainty.

Note that the minimum possible value of $\frac{J}{M}$ is roughly 0.971, so the topic signal for the platform must be extremely strong when the context signal is less than or equal to $\frac{D}{M}$ for prices to potentially decline. Note also that $\frac{D}{M}$ is a declining function in B_t, so higher values of B_t lead to lower thresholds $\frac{D}{M}$.

Predictions for short queries in Region 2 with introduction of BERT: $\gamma_q \geq \frac{D}{M}$ and $\gamma_t < \frac{J}{M}$. This means it's possible for CPC to decline if $\frac{\gamma_q^{'}}{\gamma_t^{'}}$. But, since we expect $\frac{\gamma_q^{'}}{\gamma_t^{'}}$ to be large, we expect CPC to rise for short queries in Region 2.

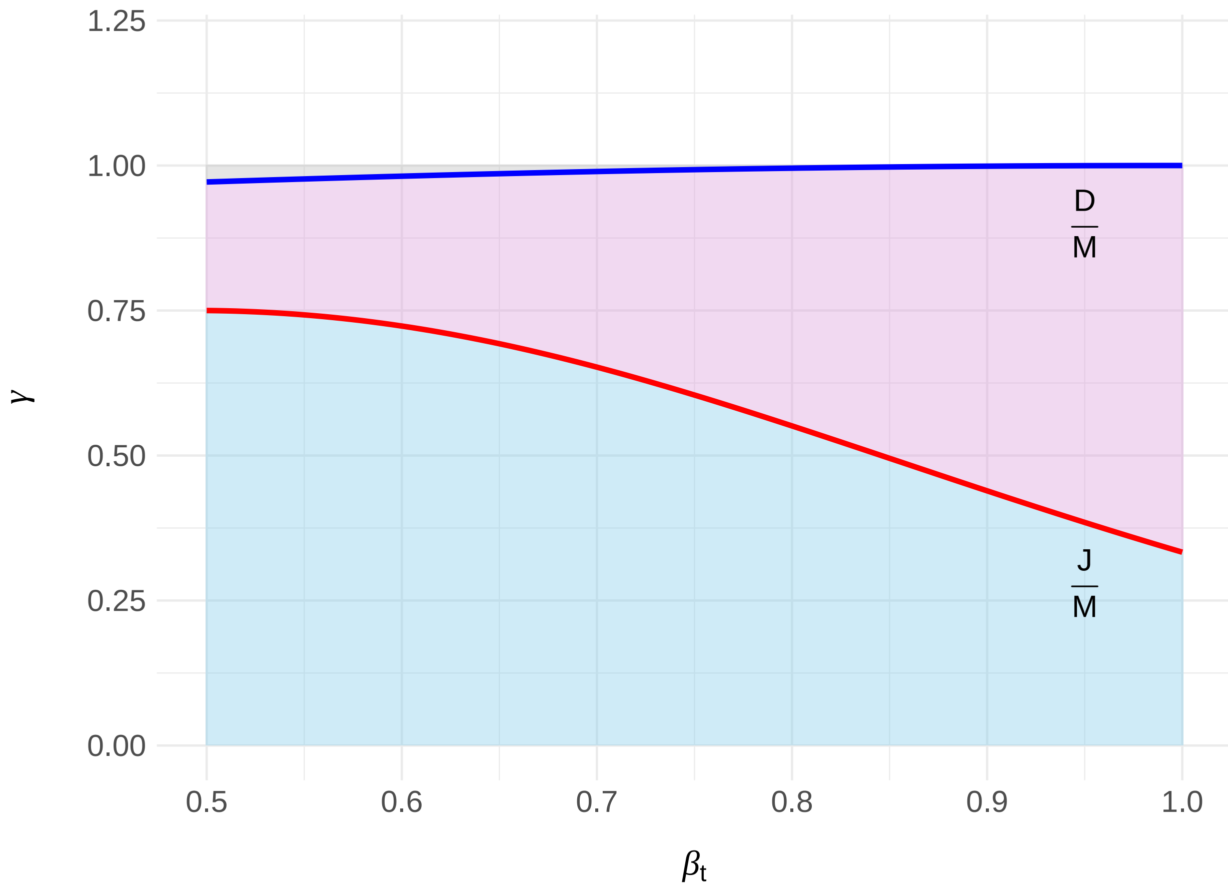
For long queries, γ_q could either be greater than or less $\frac{D}{M}$. It is unlikely $\gamma_t > \frac{J}{M}$. More likely $\gamma_t \leq \frac{J}{M}$. In either case, it is likely that prices increase.

In Figure 6 we plot the behavior of the thresholds $\frac{D}{M}$ and $\frac{J}{M}$.

Region 2: Interesting Market Outcomes We could empirically observe a decline in CPC with more bidders if the following inequality holds:

$$\frac{\gamma_q - \frac{1}{2}}{(1 - \gamma_t)} < \frac{\gamma_q^{'}}{\gamma_t^{'}} < \frac{M\gamma_q - D}{(J - M\gamma_t)}$$

Figure 6: Region 2: Threshold functions for γ_q and γ_t that impact whether prices can increase or decrease.



Alternatively, we could empirically observe an increase in CPC despite fewer bidders if the inequality flips:

$$\frac{\gamma_q - \frac{1}{2}}{(1-\gamma_t)} > \frac{\gamma_q^{'}}{\gamma_t^{'}} > \frac{M\gamma_q - D}{(J - M\gamma_t)}$$

A.4.3 Region 3

Number of Bidders The possibility for the number of bidders to decline when in Region 3 depends on if $\gamma_q \leq \frac{2}{3}$ and $\gamma_t > \frac{2}{3}$ or $\gamma_q > \frac{2}{3}$. This can be seen with the following inequality:

$$(1 - \frac{3\gamma_t}{2})\frac{\gamma_q^{'}}{\gamma_t^{'}} < \frac{3}{2}\gamma_q - 1$$

The intuition for the result can be seen by looking at the expected number of bidders $4(\gamma_t + \gamma_q) - 6\gamma_t\gamma_q$. When one signal is strong (topic or context), while the other is weak, the platform is generally running auctions with all bidders. However, when both signals are strong, the platform is predominantly allocating only the two topically relevant bidders. The number of bidders can decline when signal gains for one dimension are large while the other dimension also has a strong signal. In a sense, the platform is moving towards a more segmented market once it has strong signals of both.

Cost-Per-Click (CPC) In Region 3, the platform allocates only topically relevant advertisers when it knows the topic and context and all advertisers when it knows either just the topic or just the context. No advertiser is allocated when there is no information. Define $J = \frac{40-60B_t+45B_t^2-16B_t^3}{30(2-B_t)^2}$, $X = \frac{9+18B_t-3B_t^2+16B_t^3}{30(1+B_t)^2}$, and $Z = J + X - \frac{B(3-B)}{6}$. X is the expected CPC when allocating all four bidders when the platform knows the topic and J is the expected CPC when the platform allocates all four bidders and it only knows the context dimension. Expected CPC is therefore

$$J\gamma_q + X\gamma_t - Z\gamma_t\gamma_q$$

Taking the derivative w.r.t. a and setting it less than 0, we reach the following inequality:

$$\frac{\gamma_q^{'}}{\gamma_t^{'}}(J - Z\gamma_t) < Z\gamma_q - X$$

This inequality follows a similar structure to Region 2, except Z is different than the M we defined for Region 2.

Z is going to be equal to:

$$Z = \frac{9 + 3B_t - 28B_t^2 + 11B_t^3 + 5B_t^4}{30(1 + B_t)^2} + \frac{40 - 60B_t + 45B_t^2 - 16B_t^3}{30(2 - B_t)^2}$$

which simplifies to:

$$Z = \frac{76 - 4B_t - 150B_t^2 + 173B_t^3 - 39B_t^4 - 25B_t^5 + 5B_t^6}{30(2 - B_t)^2(1 + B_t)^2}$$

The ratio $\frac{X}{Z}$ will be equal to

$$\frac{X}{Z} = \frac{(9 + 18B_t - 3B_t^2 + 16B_t^3)(2 - B_t)^2}{76 - 4B_t - 150B_t^2 + 173B_t^3 - 39B_t^4 - 25B_t^5 + 5B_t^6}$$

The ratio $\frac{J}{Z}$ will be equal to

$$\frac{J}{Z} = \frac{(40 - 60B_t + 45B_t^2 - 16B_t^3)(1 + B_t)^2}{76 - 4B_t - 150B_t^2 + 173B_t^3 - 39B_t^4 - 25B_t^5 + 5B_t^6}$$

Under what conditions will the inequality $\frac{\gamma_q^{'}}{\gamma_t^{'}}(J - Z\gamma_t) < Z\gamma_q - X$ hold

1. $\gamma_q < \frac{X}{Z}$ and $\gamma_t \leq \frac{J}{Z}$. Inequality cannot hold and CPC must increase.

2. $\gamma_q < \frac{X}{Z}$ and $\gamma_t > \frac{J}{Z}$. Inequality can hold and prices decline if $\frac{\gamma_q^{'}}{\gamma_t^{'}}$ is large.

3. $\gamma_q > \frac{X}{Z}$ and $\gamma_t < \frac{J}{Z}$. Inequality can hold. CPC can decline if $\frac{\gamma_q^{'}}{\gamma_t}$ is small.

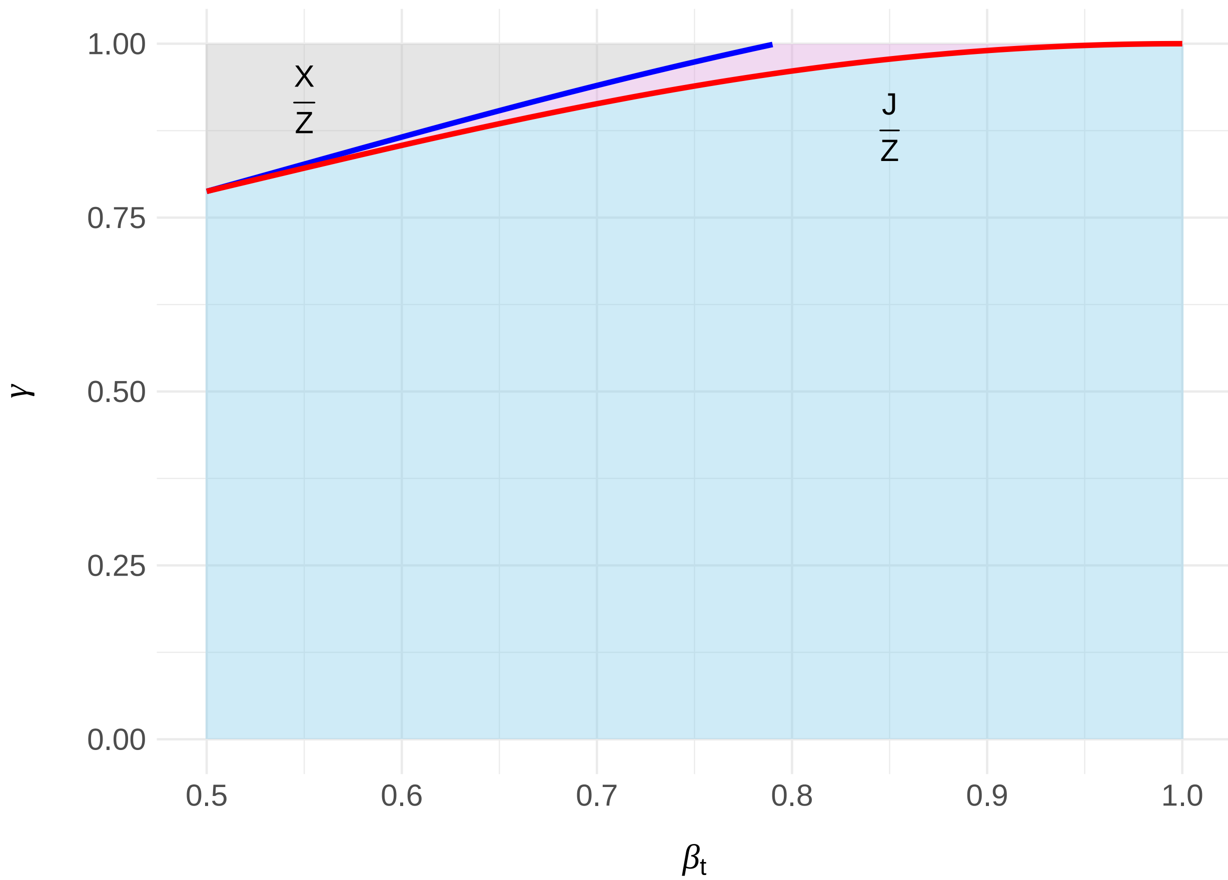
4. $\gamma_q > \frac{X}{Z}$ and $\gamma_t = \frac{J}{Z}$. Inequality holds and CPC must decline. When $\gamma_q = \frac{X}{Z}$ and $\gamma_t = \frac{J}{Z}$, the sides equal and implies there is no change to CPC.

5. $\gamma_q \geq \frac{X}{Z}$ and $\gamma_t > \frac{J}{Z}$.Inequality always holds and CPC will drop.

The behavior of CPC in Region 3 is similar to Region 2, except the thresholds are different. In Figure 7 we present the behavior of $\frac{X}{Z}$ and $\frac{J}{Z}$.

Here, the region in which γ_q and γ_t are under $\frac{X}{Z}$ and $\frac{J}{Z}$ respectively is much larger than in Region 2. The minimum value for both functions is $\frac{308}{391}$ or roughly 0.7877. If γ_q and γ_t are both less than that value, then prices cannot decline in Region 3. If we believe $\gamma_t > \frac{J}{Z}$ and $\gamma_q < \frac{X}{Z}$ and $\frac{\gamma_q^{'}}{\gamma_t}$ is large then prices are likely to decline. This may occur for longer queries with smaller B_t.

Figure 7: Region 3: Threshold functions for γ_q and γ_t that impact whether prices can increase or decrease.



Region 3: Interesting Market Outcomes We could empirically observe a decline in CPC with more bidders if the following inequality holds:

$$\frac{\frac{3}{2}\gamma_q - 1}{1 - \frac{3\gamma_t}{2}} < \frac{\gamma_q^{'}}{\gamma_t^{'}} < \frac{Z\gamma_q - X}{(J - Z\gamma_t)}$$

Alternatively, we could empirically observe an increase in CPC despite fewer bidders if the inequality flips:

$$\frac{\frac{3}{2}\gamma_q - 1}{1 - \frac{3\gamma_t}{2}} > \frac{\gamma_q^{'}}{\gamma_t^{'}} > \frac{Z\gamma_q - X}{(J - Z\gamma_t)}$$

A.4.4 Region 4

Number of Bidders Region 4 behaves similarly to Region 3, except the threshold values are higher. Now, $\gamma_q \leq \frac{4}{5}$ and $\gamma_t > \frac{4}{5}$ or $\gamma_q > \frac{4}{5}$ can lead to a decline in the number of bidders. Otherwise, the number of bidders is increasing when the inequality is flipped and 0 at equality.

$$(1 - \frac{5\gamma_t}{4})\frac{\gamma_q^{'}}{\gamma_t^{'}} < \frac{5}{4}\gamma_q - 1$$

In this region, the likelihood of seeing a decline in the number of bidders is smaller because the number of bidders is generally higher due to three bidders being allocated with full information instead of only two.

Cost-Per-Click (CPC) Region 4 differs from Region 3 because the platform is willing to allocate the three advertisers with non-zero relevancy scores to the auction, not just the topically relevant advertisers. Like Region 3, define $J = \frac{40-60B_t+45B_t^2-16B_t^3}{30(2-B_t)^2}$ and $X = \frac{9+18B_t-3B_t^2+16B_t^3}{30(1+B_t)^2}$. X is the expected CPC when allocating all four bidders when the platform knows the topic and J is the expected CPC when the platform allocates all four bidders and it only knows the context dimension. Finally, define $P = \frac{-5+20B_t-25b_t^2+15B_t^3-3b_t^4}{6B_t}$ and $L = X + J - P$. P is the expected CPC when three bidders are allocated to the auction when the platform has full information. CPC falls when:

$$\frac{\gamma_q'}{\gamma_t'}(J - L\gamma_t) < L\gamma_q - X$$

This is the same structure as Region 3, except L replaces Z.

Under what conditions will the inequality $\frac{\gamma_q'}{\gamma_t'}(J - L\gamma_t) < L\gamma_q - X$ hold

1. $\gamma_q < \frac{X}{L}$ and $\gamma_t \leq \frac{J}{L}$. Inequality cannot hold and CPC must increase.
2. $\gamma_q < \frac{X}{L}$ and $\gamma_t > \frac{J}{L}$. Inequality can hold and prices decline if $\frac{\gamma_q'}{\gamma_t'}$ is large.
3. $\gamma_q > \frac{X}{L}$ and $\gamma_t < \frac{J}{L}$. Inequality can hold. CPC can decline if $\frac{\gamma_q'}{\gamma_t'}$ is small.
4. $\gamma_q > \frac{X}{L}$ and $\gamma_t = \frac{J}{L}$. Inequality holds and CPC must decline. When $\gamma_q = \frac{X}{L}$ and $\gamma_t = \frac{J}{L}$, the sides equal and implies there is no change to CPC.
5. $\gamma_q \geq \frac{X}{L}$ and $\gamma_t > \frac{J}{L}$.Inequality always holds and CPC will drop.

L is equal to:

$$L = \frac{100 - 224B_t + 81B_t^2 + 340B_t^3 - 282B_t^4 - 119B_t^5 + 230B_t^6 - 105B_t^7 + 15B_t^8}{30(2 - B_t)^2 B_t(1 + B_t)^2}$$

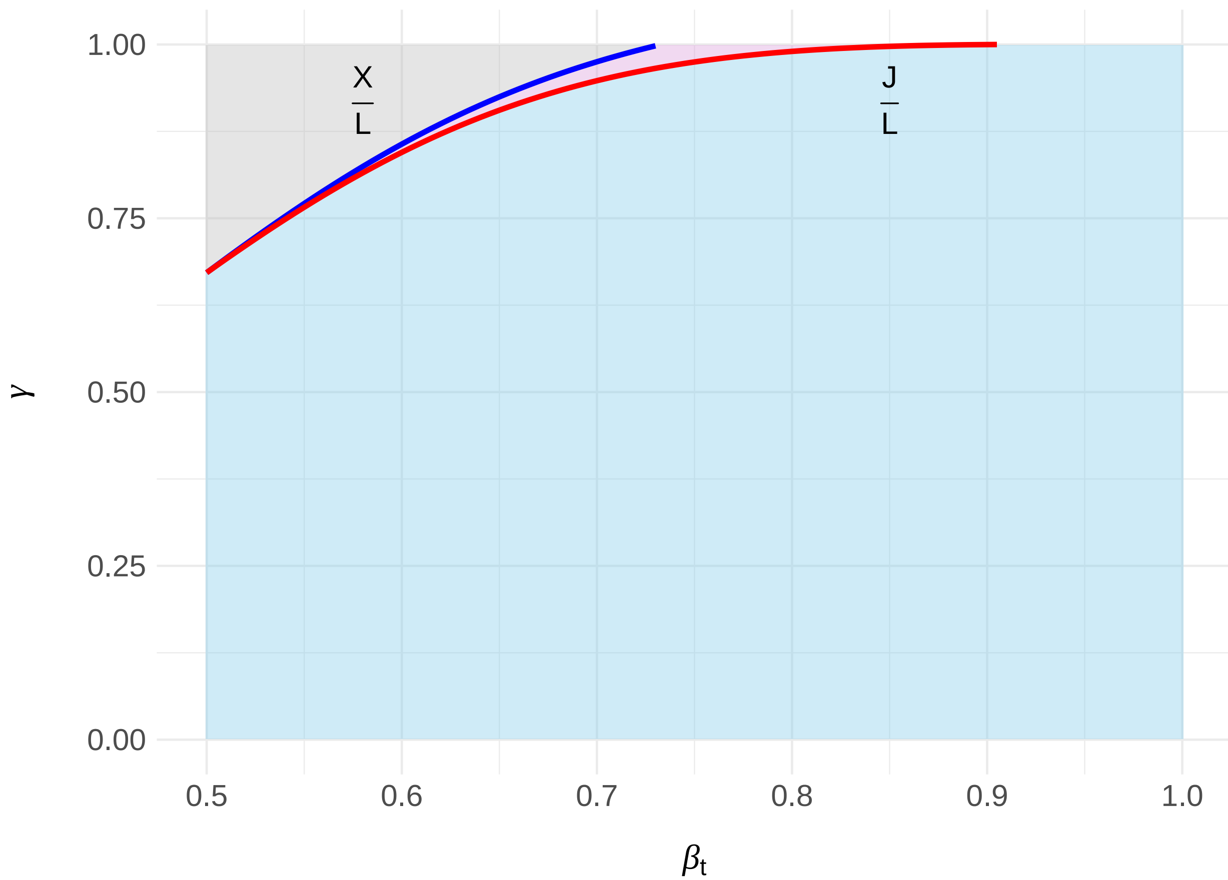
The ratio $\frac{J}{L}$ is equal to

$$\frac{J}{L} = \frac{(B_t(1 + B_t)^2)(40 - 60B_t + 45B_t^2 - 16B_t^3)}{100 - 224B_t + 81B_t^2 + 340B_t^3 - 282B_t^4 - 119B_t^5 + 230B_t^6 - 105B_t^7 + 15B_t^8}$$

and the ratio $\frac{X}{L}$ is equal to

$$\frac{X}{L} = \frac{(9 + 18B_t - 3B_t^2 + 16B_t^3)(B_t(2 - B_t)^2)}{100 - 224B_t + 81B_t^2 + 340B_t^3 - 282B_t^4 - 119B_t^5 + 230B_t^6 - 105B_t^7 + 15B_t^8}$$

Figure 8: Region 4: Threshold functions for γ_q and γ_t that impact whether prices can increase or decrease.



In Figure 8 we plot $\frac{X}{L}$ and $\frac{J}{L}$. Note that the gray area is slightly largely compared to Region 3.

Region 4: Interesting Market Outcomes We could empirically observe a decline in CPC with more bidders if the following inequality holds:

$$\frac{\frac{5}{4}\gamma_q - 1}{1 - \frac{5\gamma_t}{4}} < \frac{\gamma_q^{'}}{\gamma_t^{'}} < \frac{L\gamma_q - X}{(J - L\gamma_t)}$$

Alternatively, we could empirically observe an increase in CPC despite fewer bidders if the inequality flips:

$$\frac{\frac{5}{4}\gamma_q - 1}{1 - \frac{5\gamma_t}{4}} > \frac{\gamma_q^{'}}{\gamma_t^{'}} > \frac{L\gamma_q - X}{(J - L\gamma_t)}$$

A.4.5 Regions 5 and 6

Number of Bidders These regions see the number of bidders declining due to the platform allocating fewer bidders when it has full information. The number of bidders declines with a. The decline is faster in Region 5 compared to Region 6.

Cost-Per-Click (CPC): Region 5 Under Region 5, it is optimal for the platform to allocate all bidders except when it has full information. With full information, it allocates only the topically relevant bidders (2). Region 5 behaves similarly to Region 3. Define $F = \frac{3}{10}$. This is the average CPC when the platform allocates all bidders when it has no information. Again, define $J = \frac{40-60B_t+45B_t^2-16B_t^3}{30(2-B_t)^2}$, $X = \frac{9+18B_t-3B_t^2+16B_t^3}{30(1+B_t)^2}$, and $Z = J + X - \frac{B(3-B)}{6}$. Prices can decline if the following inequality holds:

$$\frac{\gamma_q^{'}}{\gamma_t^{'}}(J - Z\gamma_t - F(1+\gamma_t)) < Z\gamma_q - X + F(1-\gamma_q)$$

This is similar to the inequality in Region 3, except we have this extra $-F(1+\gamma_t)$ on the left-hand side and $F(1-\gamma_q)$ on the right-hand side. Ultimately, we find that critical inequality thresholds are now $\frac{J-F}{Z-F}$ for γ_t and $\frac{X-F}{Z-F}$ for γ_q. These are the same critical thresholds in Region 3, except both the numerator and denominator are reduced by F.

Note that the max value of J is $\frac{3}{10}$ when $B_t = 1$. Therefore $\frac{J-F}{Z-F} \leq 0$, meaning $\gamma_t \geq \frac{J-F}{Z-F}$ will always hold. When $B_t \leq 0.75$, $\gamma_q \geq \frac{X-F}{Z-F}$ because $\frac{X-F}{Z-F} \leq 0$. Once $B_t > 0.75$, γ_q can potentially be less than $\frac{X-F}{Z-F}$, though it must be extremely small.

In general, prices are going to just decline in Region 5 with the decline in the number of bidders. This is driven by the market segmentation effects of context.

Cost-Per-Click (CPC): Region 6 The relevant threshold inequality for Region 6 is similar to Region 4 and Region 5. Again, define $J = \frac{40-60B_t+45B_t^2-16B_t^3}{30(2-B_t)^2}$, $X = \frac{9+18B_t-3B_t^2+16B_t^3}{30(1+B_t)^2}$, $P = \frac{-5+20B_t-25b_t^2+15B_t^3-3b_t^4}{6B_t}$, and $L = X + J - P$.

$$\frac{\gamma_t^{'}}{\gamma_q^{'}}(J - L\gamma_t - F(1+\gamma_t)) < L\gamma_q - X + F(1-\gamma_q)$$

Now, the relevant thresholds are $\frac{J-F}{L-F}$ for γ_t and $\frac{X-F}{L-F}$ for γ_q. Whether prices decline or increase follow the same constraints as Region 4, except the thresholds are now $\frac{J-F}{L-F}$ and $\frac{X-F}{L-F}$. Like Region 5, $\gamma_t \geq \frac{J-F}{Z-F}$ will always hold and when $B_t \leq 0.75$, $\gamma_q \geq \frac{X-F}{Z-F}$ because $\frac{X-F}{Z-F} \leq 0$. Once $B_t > 0.75$, γ_q can potentially be less than $\frac{X-F}{Z-F}$, though it must be extremely small.

B Sampling Procedure

Below is the list of search topics we asked for in our Amazon MTurk survey.

Insurance, Travel, Home Renovation, Cars, Restaurants, How-to's, Things To Do, Company Research, Product Research, Financial Resources, Financial Products, News, Politics, Animals, Books, Movies, Shows, Clothes, Electronics, Pest Control, Birthday Gifts, Kitchen Shopping, Moving Services, Social Media, Entertainment, Cooking, Transportation, Home Decor, Health, Medicine, Cosmetics.

Each participant was randomly asked about only five of these categories.

C Alternative Identification Strategy

The main identification strategy discussed in the paper will be biased if there exists time-varying unobservables correlated with the introduction of BERT and auction outcomes. Given our findings, such a confound would need to be a significant event that uniformly drives up the number of bidders across queries and differentially affects CPC across query lengths. While such a confounder seems unlikely, as in any observational study, we cannot completely rule it out.

To strengthen the validity of our results, we present an alternative identification strategy that exploits the fact that BERT is likely to affect queries differently due to their inherent linguistic properties. In doing so, we compare queries more likely to be affected by BERT with queries less likely to be affected by BERT before and after the introduction of BERT. Under this identification strategy, a confound must be year-month specific *and* correlate with query linguistic properties.

C.1 Defining Query Metrics

The Computer Science literature has documented that BERT better understands complex syntactic language structures and semantic relationships between words and sentences (Devlin et al., 2018; Lin et al., 2019; Tenney et al., 2019a,b; Rogers et al., 2021). A complex query will benefit from BERT's ability to capture syntactic information, while all queries, including those with simple syntactic structures, will experience changes to semantic relationships and understanding.

Motivated by these observations, we create two measurement variables, Linguistic Complexity and Cosine, to capture query syntactic complexity and semantics. Under the assumption that the interpretation and use of these linguistic properties changed with the introduction of BERT, we can identify variation in our dependent variables caused by BERT's implementation interacting with our measurements.

Linguistic Complexity BERT can better understand complex syntactic language structures (Lin et al., 2019; Tenney et al., 2019a). Therefore, BERT's introduction should change how Google handles complex syntactic language. Examples of complex syntactic information include query syntax tree structures (Lin et al., 2019), parts of speech (Tenney et al., 2019a), and sentence dependency features Tenney et al. (2019a); Liu et al. (2019). Hewitt and Manning (2019) finds that syntax trees, i.e., hierarchical characterizations of grammatical language structures, are embedded in BERT vector spaces and Tenney et al. (2019b) finds that Parts of Speech (POS) tags are also present in BERT vector spaces. These findings help us identify the information BERT will interact with and drive the design of our first measurement: Linguistic Complexity (LC).

For each query, we measure the depth of the syntax tree and count the unique number of Parts Of Speech (POS). LC is defined as a dummy variable and takes on a value of 1 if either a query's syntax tree depth is greater than the median depth in our dataset (median = 2) or the unique POS count is greater than the median county in our dataset (median = 2). Otherwise, it's 0. LC captures the syntactic complexity of the query.

Cosine BERT understands semantics and how words and concepts *relate* to each other (Tenney et al., 2019b). It knows that socks relate to shoes, banks can relate to bodies of water or financial institutions, and computers can sometimes relate to mice. This semantic knowledge improves Google's ability to interpret and categorize search queries, ultimately affecting the relationships between queries. This latter observation is critical as changes to query relations likely impact Google's matching process and identification of relevant advertisers. Understanding that a query relates to more (fewer) queries will likely lead to more (fewer) advertisers being deemed relevant.These observations motivate the design of our second linguistic property: Cosine.

At a high level, we measure semantic changes (Cosine) using the difference in the number of queries related to a given query before and after BERT.

To define Cosine, we must first identify the primary interpretation algorithm used by Google before BERT (RankBrain). While we cannot be certain, RankBrain is likely Doc2Vec (D2V) or Word2Vec (W2V), the former being a more generalized form of W2V.[25]

For a query i, we generate a vector representation with BERT and D2V models. We then calculate the cosine similarity score between query i and all other queries in the dataset *within* a given model vector space. This generates two $N \times N$ cosine similarity matrices, one for the BERT vector space and one for the D2V vector space, where N is the number of queries in our dataset and matrix element i, j is the cosine similarity score between query i and query j.

We use these matrices to measure changes in query semantic relationships across D2V and BERT vector spaces. Specifically, we measure the set of queries that pass "relatedness" thresholds *within* vector spaces and then compare differences in these sets *across* vector spaces. We first calculate each cosine matrix's 25th, 50th, and 75th percentile cosine scores. These are our "relatedness" thresholds.[26] Then, for each query i, we identify the set of other queries with a cosine similarity score greater than or equal to the chosen threshold (i.e., 25, 50, or 75) in the respective vector spaces. For example, if the 75th threshold in the D2V space is 0.5 and the query "shoes" has a cosine similarity score of 0.8 with the query "socks" and 0.2 for the query "hat", then the "socks" query will be in the relevant set for the D2V space. If, in the BERT space, the 75th threshold is 0.8 and the cosine score between "shoes" and "hats" is 0.7, then "hat" would now appear in the relevancy set for "shoes" in the BERT space. This gives us two sets of relevant queries per query, one for each vector space.

Then, we calculate the differences between the relevancy sets across BERT and D2V for a given query. We define "added" queries as those that did not appear in the D2V relevancy

[25]Google filed patents: `https://patents.google.com/patent/US9740680B1/en`. See also: `https://opensource.googleblog.com/2013/08/learning-meaning-behind-words.html` for technology before RankBrain that looks similar to Word2Vec, and the researchers for both Word2Vec and Doc2Vec were employed at Google.

[26]We vary the thresholds to generalize the measurement. We focus on thresholds because we reason that Google likely has some cut-off requirement determining whether an advertiser is relevant enough to a given query.

set but did appear in the BERT set. Alternatively, we define "removed" queries as those that appeared in the D2V set but did not appear in the BERT set. For a given query i, we sum up the total number of "added" and "removed" queries and take the ratio between the two sums. We define Cosine as this ratio.

A ratio value equal to one tells us that just as many queries were added as removed. Ratio values less than one indicate that more queries were removed than added, indicating that BERT believes the query is more specific and related to fewer other queries. Finally, ratio values greater than 1 tell us that more queries were added than removed, suggesting that BERT believes the query is related to more queries. Because the ratio of added to removed queries is skewed to the right (median = 1.094, mean = 175.605), we use the log of 1 + Cosine in our model.[27]

It is worth noting that we use this process to generate Cosine because the D2V and BERT vector spaces and their cosine scores are not directly comparable without making strong assumptions about what the dimensions of each model's vector space represent. D2V and BERT capture potentially different sets of information, making vector comparison and projection methods infeasible. Additionally, the scales of these vector spaces are potentially different, meaning we cannot directly compare query cosine scores across algorithms. Therefore, to effectively measure query semantic changes, we must adequately standardize measurements *within* vector spaces such that they can then be compared *across* vector spaces.

Average Query Metrics Table 12 and 13 present average Cosine and LC measures, respectively, for all queries ("All") and by keyword length. Table 13 also presents the proportion of queries with POS counts and syntax tree depths greater than the median. (A value of 1 means above the median for the particular measurement). We identify several patterns. First, LC measures correlate with query length due to the presence of within-query contextual information, inherently making it longer. Second, short queries have, on average,

[27] In Appendix **??**, we show that our results are robust to how we define Cosine.

a large Cosine score, while long queries have a low Cosine score. The average long query ratio between "added" and "removed" is 0.78, indicating that BERT finds these queries more specific than D2V.

Table 12: Average Log(Cosine+1).

Keyword Length	75th Ratio	50th Ratio	25th Ratio
All	1.761	1.324	1.049
Short	3.107	2.348	1.645
Medium	1.393	1.043	0.914
Long	0.650	0.481	0.451

These values are the average log(Cosine + 1), where Cosine is the ratio between added and removed sets.

Table 13: Average Linguistic Complexity.

Keyword Length	Linguistic Complexity	Tree Height	POS
All	0.613	0.566	0.400
Short	0.004	0.001	0.004
Medium	0.811	0.730	0.439
Long	1.000	0.998	0.986

C.2 Specification

Given the two linguistic metrics defined above, we estimate the following model:

$$Y_{i,t} = \beta_1 Post_t \times LC_i + \beta_2 Post_t \times log(Cos_i) + \beta_3 Post_t \times LC_i \times log(Cos_i) + \delta_i + \gamma_t + \epsilon_{i,t}, \quad (6)$$

where *Post* takes on a value of 1 for the months post-BERT (November to February), and 0 otherwise. LC_i is Language Complexity of query i and Cos_i is the the Cosine of query i. δ_i are query fixed effects and γ_t year-month fixed effects. β_1 captures the variation in Y that is explained by the effect of LC after BERT gets introduced, β_2 captures the variation in Y explained by the effect of changes to Cosine after the introduction of BERT, and β_3 captures the interaction between LC and Cos after the introduction of BERT.

We include query and year-month fixed effects, δ_i and γ_t, respectively. We estimate Equation 6 using data from July 2019 to February 2020 and cluster standard errors at the query level. To replicate the findings in the previous section, we restrict this analysis to competitive queries, i.e., those queries with non-zero mode CS scores from July to October 2019.

Identification Assumptions This identification strategy relies on the assumptions that: (1) the linguistic properties we defined affect CPC and CS only through their interaction with the query interpretation system; (2) BERT interacts with our linguistic measurements differently than the previous algorithm; (3) queries with high and low values of Linguistic Complexity and Cosine are comparable; (4) there are no time-varying-query linguistic type-specific confounders.[28]

To partially validate these assumptions, we perform three tests. To test assumption three, in Appendix **??**, we use an event study-like model to show that we estimate null pre-trends. To account for consumer search behavior, in Appendix **??**, we show that results hold when we control search volume. Finally, in Section **??**, we perform a placebo test that shows that our estimates are likely due to the introduction of BERT (partially validating assumptions one and two).

C.3 Results

We present the estimates of Equation 6 using CS as the dependent variable in Table 14. $Post_t \times LC_i$ is positive and significant, suggesting that when Linguistic Complexity is above the median, CS increases by $0.33 - 0.4\%$. These findings suggest that when syntactic information acquisition opportunities are significant, BERT's information increases Google's

[28]For example, changes in search behavior that are query-type specific and unrelated to BERT can be problematic. Consider the case in which wealthier consumers begin submitting search queries that are more linguistically complex because of something other than BERT. If advertisers change their targeting strategies and increase their bids in response to this behavior, our results will be upward biased.

Table 14: Alternative Strategy: $log(CS)$

	(1) 75th	(2) 50th	(3) 25th
Post × Linguistic Complexity	0.0040**	0.0033*	0.0033*
	(0.002)	(0.002)	(0.002)
Post × Log(Cosine)	0.0012**	0.0009*	0.0007
	(0.001)	(0.001)	(0.001)
Post × Linguistic Complexity × Log(Cosine)	−0.0015**	−0.0020**	−0.0030**
	(0.001)	(0.001)	(0.001)
Observations	63,474	63,474	63,474
R^2	0.9786	0.9785	0.9785

Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: Regressions include year-month and query fixed effects. Standard errors clustered at the query level are reported in parentheses. Data from July 2019 to Feb 2020.

confidence that it correctly interprets the query and increases auction density. Second, consistent with our expectations, $Post_t \times Cos_i$ is also positive and significant. The coefficient estimates suggest that a 1% increase Cosine translates to about a 0.1% increase in CS. Finally, as expected, the interaction term between LC and Cosine is negative. Linguistically complex queries are generally more specific, and BERT learns this, leading to partial market filtering and some removal of advertisers.

To put these results into context, we can use the average keyword length values in Table 12 and Table 13 to bootstrap predicted changes to the average number of bidders. We find that aggregate CS increases by a conservative $0.1 - 0.3\%$. When we predict average values by keyword length, we find that, across lengths, CS is increasing by about $0.1 - 0.4\%$ (See Table 15). The uniform increase in CS is consistent with our empirical observations using the main DD setup. All estimates are more conservative, likely due to measurement errors in our treatment variables.

In Table 16, we present the results for $log(CPC)$. Consistent with our expectations, a 1%

Table 15: Predicted $log(CS)$ estimates by model specification.

Keyword Length	75th	50th	25th
All	0.003***	0.002***	0.001***
Short	0.004***	0.002***	0.001***
Medium	0.003***	0.002***	0.001***
Long	0.004***	0.003***	0.002***

Note: 99% confidence intervals are estimated by bootstrapping predictions (1000 iterations). ***p<0.01.

increase in Cosine positively increases CPC by about 0.6%. Inconsistent with our predictions, we find that $Post_t \times LC_i$ is negative and significant. Finally, the triple interaction is close to zero and not significant.

Table 16: Alternative Identification: $log(CPC)$

	(1) 75th	(2) 50th	(3) 25th
Post × Linguistic Complexity	−0.0270*	−0.0289**	−0.0307**
	(0.015)	(0.014)	(0.015)
Post × Log(Cosine)	0.0056*	0.0065*	0.0075
	(0.003)	(0.004)	(0.005)
Post × Linguistic Complexity × Log(Cosine)	−0.00003	0.0009	0.0006
	(0.005)	(0.006)	(0.009)
Observations	63,474	63,474	63,474
R^2	0.795	0.795	0.795

Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: Regressions include year-month and query fixed effects. Standard errors clustered at the query level are reported in parentheses. Data from July 2019 to Feb 2020.

We again use the average values in Tables 12 and 13 to put these results into context. We present the estimated effect on CPC in Table 17. We find that the average CPC decreases by roughly 0.8%. When we break it down by keyword length, we find that CPC increases for short queries by roughly 1.5%, medium queries see CPC decrease by about 1.6%, and long queries see CPC decrease by approximately 2.5%. These average values are broadly consistent with our primary DD analysis, though we note that average CPC declines under this specification and is null in our main DD result.

Table 17: Predicted $log(CPC)$ estimates by model specification.

Keyword Length	75th	50th	25th
All	-0.007^{***}	-0.008^{***}	-0.011^{***}
Short	0.017^{***}	0.015^{***}	0.012^{***}
Medium	-0.014^{***}	-0.016^{***}	-0.018^{***}
Long	-0.023^{***}	-0.025^{***}	-0.027^{***}

Note: 99% confidence intervals are estimated by bootstrapping predictions (1000 iterations). ***p<0.01.

C.4 Robustness Checks

Placebo Test Our identification rests on the assumption that the effects we observe are due to the linguistic variables interacting with a change in the interpretation algorithm (i.e., BERT). We support this assumption by performing a placebo test. We estimate 6 using the same period but a year before (July 2018 to February 2019) and create a placebo post-BERT variable that takes on a value of 1 for November through February, 0 otherwise. Since no known change exists that we suspect interacts with these measurements during this time frame, we expect to estimate insignificant results. In Tables 18 and 19 we present null results.

Changes in Organic Rank Results As we did for the main identification strategy in the paper, we also test whether our results hold when controlling for organic rank changes. We re-estimate Equation 6 including Featured Snippet usage, the Number of SERP features, and whether the top domain has changed compared to the previous month. We present these results in Table 20 and Table 21 for CS and CPC, respectively. We find results that are directionally consistent but attenuated and more imprecise, particularly for CPC, where

Table 18: Placebo Test: $log(CS)$.

	(1) 75th	(2) 50th	(3) 25th
Post × Linguistic Complexity	0.001	0.001	0.0004
	(0.001)	(0.001)	(0.001)
Post × Log(Cosine)	0.0003	0.0003	0.0004
	(0.0002)	(0.0002)	(0.0004)
Post × Linguistic Complexity × Log(Cosine)	−0.0002	−0.0002	−0.00001
	(0.0002)	(0.0003)	(0.001)
Observations	63,468	63,468	63,468
R^2	0.997	0.997	0.997

Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: Regressions include year-month and query fixed effects. Standard errors clustered at the query level are reported in parentheses. Data from July 2018 to Feb 2019.

they become insignificant.[29]

Changes in Consumer and Advertiser Behavior Similar to what we did for the main identification strategy, here we re-estimate our alternative identification regression models controlling for Google Search trends Interest and SEMRush's search volume. In Table 22, we present changes to CS, controlling for both.

Similarly, in Table 23, we present percent changes to CPC controlling for Interest and search volume. We find that estimates remain largely consistent with our primary specification.

[29]The loss in precision is because we do not have organic rank information for all queries (only for $41,555$ out of $63,394$). Indeed, we find that using the same subset of queries for which we have organic rank information and not including controls, we obtain very similar results to those reported in Table 20 and Table 21

Table 19: Placebo test: $log(CPC)$.

	(1) 75th	(2) 50th	(3) 25th
Post × Linguistic Complexity	0.001	0.002	0.001
	(0.002)	(0.002)	(0.002)
Post × Log(Cosine)	0.0004	0.0005	0.0004
	(0.001)	(0.001)	(0.001)
Post × Linguistic Complexity × Log(Cosine)	0.001	−0.0002	0.001
	(0.002)	(0.001)	(0.002)
Observations	63,468	63,468	63,468
R^2	0.979	0.979	0.979

Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: Regressions include year-month and query fixed effects. Standard errors clustered at the query level are reported in parentheses. Data from July 2018 to Feb 2019.

Table 20: Controlling for organic rank changes: $log(CS)$

	(1)	(2)	(3)
	75th	50th	25th
Post × Linguistic Complexity	0.0046*	0.0039*	0.0033
	(0.0024)	(0.0023)	(0.0024)
Post × Cosine	0.0011*	0.001	0.007
	(0.0006)	(0.0007)	(0.0010)
Post × Linguistic Complexity × Cosine	−0.0011	−0.0013	−0.0016
	(0.0008)	(0.0010)	(0.0014)
log(SERP Features Count)	−0.0003	−0.0003	−0.0003
	(0.0004)	(0.0004)	(0.0004)
Feature Snippet	0.0069***	0.0069***	0.0069***
	(0.0022)	(0.0022)	(0.0022)
Domain Change	−0.0008	−0.0008	−0.0008
	(0.0008)	(0.0008)	(0.0008)
Observations	41,555	41,555	41,555
R^2	0.9772	0.9772	0.9772
Note:	*p<0.1; **p<0.05; ***p<0.01		

Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: All regressions include year-month and query fixed effects. Standard errors clustered at the query level are reported in parentheses. Data is from July 2019 to February 2020.

Table 21: Controlling for organic rank changes: $log(CPC)$

	(1) 75th	(2) 50th	(3) 25th
Post × Linguistic Complexity	−0.0158	−0.0211	−0.0206
	(0.0178)	(0.0169)	(0.0176)
Post × Cosine	0.0052	0.0050	0.0065
	(0.0039)	(0.0044)	(0.0063)
Post × Linguistic Complexity × Cosine	−0.0043	−0.0022	−0.0047
	(0.0061)	(0.0075)	(0.0104)
log(SERP Features Count)	0.0113***	0.0113***	0.0113***
	(0.0032)	(0.0032)	(0.0032)
Feature Snippet	−0.0168	−0.0167	−0.0167
	(0.0133)	(0.0133)	(0.0133)
Domain Change	−0.0078	−0.0077	−0.0077
	(0.0057)	(0.0057)	(0.0057)
Observations	41,555	41,555	41,555
R^2	0.8038	0.8038	0.8038
Note:	*p<0.1; **p<0.05; ***p<0.01		

Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: All regressions include year-month and query fixed effects. Standard errors clustered at the query level are reported in parentheses. Data is from July 2019 to February 2020.

Table 22: Controlling for search interests and search volume: $log(CS)$

	(1)	(2)	(3)
	75th	50th	25th
Post × Linguistic Complexity	0.0042**	0.0036*	0.0036*
	(0.002)	(0.002)	(0.002)
Post × Cosine	0.0012**	0.0010*	0.0007
	(0.0005)	(0.0006)	(0.0008)
Post × Linguistic Complexity × Cosine	−0.0015**	−0.002**	−0.003***
	(0.0007)	(0.0008)	(0.0012)
Interest	0.00004	0.00004	0.00004
	(0.00004)	(0.00004)	(0.00004)
log(SV)	0.0062***	0.0062***	0.0062***
	(0.0023)	(0.0023)	(0.0023)
Observations	63,394	63,394	63,394
Note:			*p<0.1; **p<0.05; ***p<0.01

Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: All regressions include year-month and query fixed effects. Standard Errors clustered at the query level are reported in parentheses. Data is from July 2019 to February 2020.

Table 23: Controlling for search interests and search volume: $log(CPC)$

	(1)	(2)	(3)
	75th	50th	25th
Post × Linguistic Complexity	-0.0254^{*}	-0.0274^{**}	-0.0290^{**}
	(0.0147)	(0.0139)	(0.0145)
Post × Cosine	0.0057^{*}	0.0066^{*}	0.0077
	(0.0032)	(0.0037)	(0.0052)
Post × Linguistic Complexity × Cosine	−0.0003	0.0006	0.00001
	(0.0051)	(0.0064)	(0.0092)
Interest	0.0001	0.0001	0.0001
	(0.0003)	(0.0003)	(0.0003)
log(SV)	0.0142	0.0142	0.0141
	(0.0129)	(0.0129)	(0.0129)
Observations	63,394	63,394	63,394
R^2	0.7951	0.7951	0.7950
Note:			*p<0.1; **p<0.05; ***p<0.01

Significance Levels: *p<0.1; **p<0.05; ***p<0.01.
Note: All regressions include year-month and query fixed effects. Standard Errors clustered at the query level are reported in parentheses. Data is from July 2019 to February 2020.