TI 2019-056/III
Tinbergen Institute Discussion Paper

Dynamics in clickthrough and conversion probabilities of paid search advertisements

Anoek Castelein[1]
Dennis Fok[1]
Richard Paap[1]

[1] Econometric Institute, Erasmus University Rotterdam

Dynamics in clickthrough and conversion probabilities of paid search advertisements*

Anoek Castelein[a] Dennis Fok[a] Richard Paap[a]

[a] Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam

August, 2019

Abstract

We develop a dynamic Bayesian model for clickthrough and conversion probabilities of paid search advertisements. These probabilities are subject to changes over time, due to e.g. changing consumer tastes or new product launches. Yet, there is little empirical research on these dynamics. Gaining insight into the dynamics is crucial for advertisers to develop effective search engine advertising (SEA) strategies. Our model deals with dynamic SEA environments for a large number of keywords: it allows for time-varying parameters, seasonality, data sparsity and position endogeneity. The model also discriminates between transitory and permanent dynamics. Especially for the latter case, dynamic SEA strategies are required for long-term profitability.

We illustrate our model using a 2 year dataset of a Dutch laptop selling retailer. We find persistent time variation in clickthrough and conversion probabilities. The implications of our approach are threefold. First, advertisers can use it to obtain accurate daily estimates of clickthrough and conversion probabilities of individual ads to set bids and adjust text ads and landing pages. Second, advertisers can examine the extent of dynamics in their SEA environment, to determine how often their SEA strategy should be revised. Finally, advertisers can track ad performances to timely identify when keywords' performances change.

Key words: Clickthrough, Conversion, Search engine advertising, Dynamic, Endogeneity, Time-varying parameters, Bayesian.

*We thank Rutger-Jan Lange and participants at the 46th European Marketing Academy Conference (Groningen, 2017), the 14th Marketing Dynamics Conference (Hong Kong, 2017), the 5th Econometric Institute PhD Conference (Rotterdam, 2017) and at the seminars of the Erasmus University Rotterdam and the University of Groningen for their helpful comments and suggestions.

1. Introduction

Search engine advertising (SEA) has become an important marketing channel for firms (Ryan, 2016). SEA is an advertising form that allows firms to place advertisements on the search results pages of search engines such as Google, Yahoo! and Bing. Search engines select the ads to be shown based on an individual's search, enabling advertisers to target individuals. Most search engines use an auction to select the ads. In general, the advertiser who sets the highest (quality adjusted) bid obtains the most prominent position. Usually, up to three ads are shown both on top and on the bottom of the search results page for a given search query.

In designing a SEA strategy, an advertiser has to create text ads and landing pages, determine the search phrases for which an ad is eligible to show up (the *keywords*), and set a bid on each keyword. When the performances of ads change over time, a SEA strategy requires regular revision to be effective. Dynamics in ad performance can result from e.g. the introduction of new products, changes in consumer tastes or populations, the entering of new competitors, or seasonality. Yet, there is little empirical research on dynamics in ad performance.

In this paper, we develop a dynamic Bayesian model for the performance of paid search ads in Google. The proposed model is especially suited to deal with dynamic SEA environments. It allows for dynamics through seasonal effects and time-varying parameters, and discriminates between permanent and transitory dynamics. Especially when shocks are long lasting, dynamic SEA strategies are required for long-term profitability. In our empirical application, we find evidence of substantial persistent time variation in ad performance, emphasizing the importance of addressing dynamics in ad performance models.

We model ad performance in terms of the clickthrough and conversion probabilities of keywords. In the context of SEA, the term "keyword" refers to a string of words, e.g. a keyword can be quite generic ("laptop") or more specific ("laptop acer vn7 571g"). An advertiser links each of her ads to a set of keywords, and sets a bid on each keyword. When a consumer's search query matches a keyword, the associated ad will be eligible for the auction. The clickthrough probability is the probability that a consumer who gets served an ad due to the keyword, clicks on the ad. The conversion probability is the probability that a consumer who has already clicked on the ad converts, that is, buys a product or service.

A number of studies have constructed models for clickthrough and conversion probabilities of keywords (see Rutz & Bucklin, 2007, Ghose & Yang, 2009, Agarwal, Hosanagar, & Smith, 2011, and Rutz, Bucklin, & Sonnier, 2012). Our study differs from these papers by

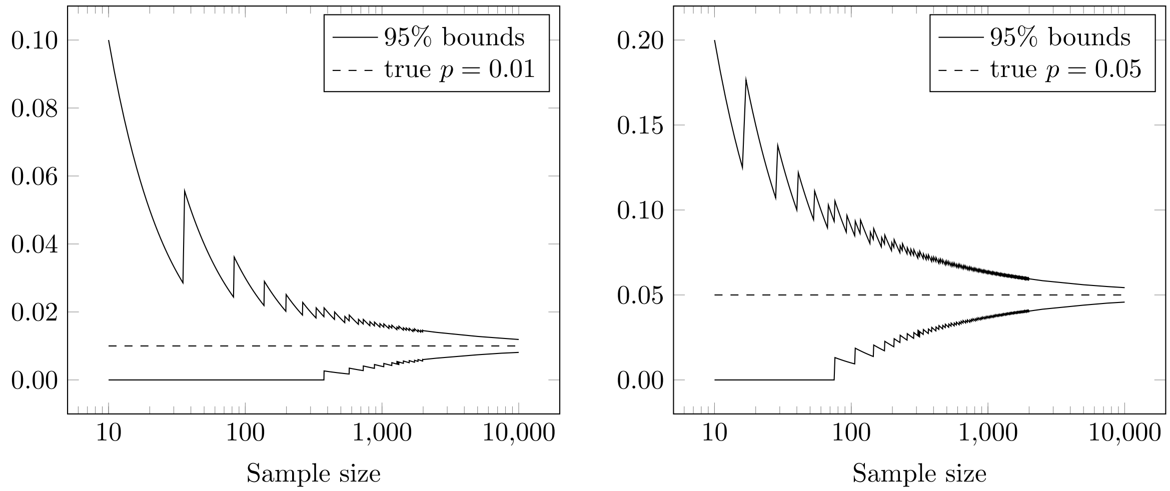


Figure 1: Conservative 95% bounds of the observed proportion of clicks/conversions as a function of the sample size on a logarithmic scale (true probability equals 1% (left) or 5% (right)).

focusing on the *dynamics* of keyword performance.

Despite the availability of advertiser-level ad performance data, a number of challenges arise when estimating clickthrough and conversion probabilities of keywords. Next to the challenge of potential dynamics in ad performance, a second challenge is data sparsity. The majority of keywords in an advertiser's portfolio generate only little traffic, that is, few users search for that keyword. For these so-called sparse keywords, taking simple averages of realized clicks and conversions in the past is insufficient to estimate the clickthrough and conversion probabilities as these estimates can be highly inaccurate. We illustrate this in Figure 1, which shows the proportion of clicks (conversions) we can expect to observe given a certain sample size (number of impressions[1] or clicks) for two realistic probabilities: 1% and 5%. The figure shows that one needs at least a couple of hundred impressions (clicks) for the sample average to be a reliable estimator of the true clickthrough (conversion) probability. If a keywords generates little traffic, it may take a long time before this number of impressions and clicks are collected. Within this time frame the true probability may, in practical situations, already have changed.

The third challenge is estimating the causal effect of ad position on ad performance, as the position is endogenously related to clickthrough and conversion probabilities. There are three sources of endogeneity:

i There is a potential reversed causality relationship due to strategic bidding behavior: an advertiser might bid more for keywords with a high expected clickthrough and conversion

[1]The number of impressions is the number of times an ad is shown on the search results page.

probability, to obtain a favorable position for these keywords.

ii There is a reversed causality relationship due to a quality adjustment in the keyword auction (Google uses the so-called quality score): position might not only affect clickthrough probabilities, but reversely, the previous clickthrough rates affect the position through their impact on the quality adjustment.

iii There is a potential confounding factor: competition is likely to affect both keyword performance and ad position, but is unobserved by the advertiser.

In this paper, we propose a dynamic model that addresses all above challenges by allowing for explained and unexplained dynamics, data sparsity, missing data, position endogeneity and unobserved heterogeneity across keywords. The model captures unexplained dynamics through time-varying parameters that follow either stationary or nonstationary AR(1) processes to distinguish between transitory and permanent dynamics. The model addresses the sparsity problem by linking keywords to each other based on common factors such as semantic keyword characteristics. Finally, the model accounts for position endogeneity in the manner proposed by Ghose and Yang (2009). That is, we simultaneously model the consumers' clickthrough and conversion behavior, the search engine's position allocating behavior, and the firm's bidding behavior, and correlate the error terms of the equations with each other.

The resulting model is estimated using a Bayesian approach. We develop an efficient Gibbs sampler with Polya-Gamma data augmentation for the logit part of the model (Geman & Geman, 1987, Tanner & Wong, 1987, Polson, Scott, & Windle, 2013) in which we draw the time-varying parameters using the forward-filtering backward-sampling algorithm of Durbin and Koopman (2002). This efficient approach is crucial to be able to use the methodology at a daily frequency for a realistically large number of keywords. It also deals naturally with missing data.

We illustrate the model using a unique dataset from a Dutch online retailer that sells laptops and advertises on Google. The data consists of the historical performance of 14,710 keywords measured at a daily frequency over the period January 2014 until March 2016. We find substantial time variation in clickthrough and conversion probabilities, indicating that a dynamic SEA strategy is required. Furthermore, we find that shocks mostly have a permanent or highly persistent effect on clickthrough probabilities; this holds for market-level shocks and most brand-level shocks. For conversion probabilities the shocks have different effects on different type of ads. Whereas market shocks permanently affect conversion probabilities, most brand-level shocks have a more transitory effect. Finally, Bayes factors indicate that the dynamic model is substantially better in forecasting ad performance than the static model.

We also find evidence of position and bidding endogeneity, indicating that purely predictive models are unable to capture causal relationships between ad position and clickthrough and conversion probabilities.

The managerial implications of this paper are threefold. First, advertisers can use the model to obtain accurate daily estimates of clickthrough and conversion probabilities of individual keywords. These estimates can be used to set bids, adjust text ads and landing pages, and to identify keywords whose performance is divergent from similar keywords. Second, advertisers can examine the extent of dynamics in their SEA environment, to determine how often their bidding strategy should be revised. In doing so, advertisers can discriminate between keywords by using the persistence and influence of shocks on different types of keywords. Finally, advertisers can use the model to track the performance of keywords to timely identify when this performance changes.

The remainder of this paper is organized as follows. In Section 2 we discuss the background for this research. We explain how the mechanism underlying SEA works and discuss related work on modeling clickthrough and conversion probabilities. In Section 3 we briefly discuss the data generally made available by search engines to further discuss the context of SEA. Section 4 is devoted to a detailed discussion of the methodology. We show empirical results in Section 5 including an analysis of the model's predictive performance against a static model. Section 6 discusses the managerial implications of this research. We conclude with a summary and a critical discussion. Finally, Appendix A documents our efficient Gibbs sampler in detail.[2]

2. Background

2.1. The mechanism underlying search engine advertising

From the search engine's perspective, much literature has focused on the mechanism design of the keyword auction (see e.g. Borgs et al., 2007, Cary et al., 2007, Edelman, Ostrovsky, & Schwarz, 2007, and Yao & Mela, 2011). The design Google and Yahoo! use is formally known as a generalized, second-price, sealed-bid auction (Edelman et al., 2007).

[2] The Supplementary Materials, containing all results of the empirical application as well as trace plots and effective sample sizes of the MCMC output, are available upon request.

This real-time keyword auction works as follows. Advertisers link their ads to keywords and place a bid on each keyword. The bid indicates the maximum amount the advertiser is willing to pay for a click. Some search engines such as Google also assign a quality score to an ad, to adjust the bids for relevance of the advertised website. Next, when a consumer enters the search query at a search engine, the engine considers all advertisers' ads for which the associated keywords match the consumer's search. The available ad slots are allocated according to the advertisers' quality adjusted bids. The search engine only charges the advertiser a fee when a consumer clicks on the ad; this fee is known as the cost-per-click (CPC). The CPC is based on the bid of the ad that is ranked just below (the second price), corrected for the quality scores of these two ads. The CPC is thus not necessarily equal to the bid, but it is never higher.

The distinct feature of the generalized, second-price auction is that bidders pay a price based on the bid of the advertiser ranked below. Hereby, search engines avoid that advertisers use cycling bidding strategies to optimize profits, that is, that advertisers continuously decrease their bids until they obtain a less prominent ad position after which they increase their bids again (Borgs et al., 2007).

2.2. Modeling clickthrough and conversion probabilities of keywords

From the advertisers' perspective, some literature has focused on modeling clickthrough and conversion probabilities of keywords (see e.g. Ghose & Yang, 2009, Agarwal et al., 2011, and Rutz et al., 2012). The key focus of these studies is addressing position endogeneity.

To better understand the sources of position endogeneity, we conceptualize the mechanism underlying the keyword auction in Figure 2. Based on the inputs of the keyword auction (the advertiser's bid and quality score and the competitors' bids and quality scores) the search engine determines the ad position and the cost-per-click. The ad position then potentially affects consumers' clickthrough and conversion behavior.

There are three potential sources of endogeneity. First, competition can be a confounding factor as it is unobserved and both enters the keyword auction to determine the ad position as well as potentially affects clickthrough and conversion probabilities through consumer behavior. Second, there is a potential reversed causality problem as some search engines, including Google, use the past clickthrough rates to assign quality scores to keywords to determine the ad position. Finally, there is a second potential reversed causality problem due to strategic bidding behavior. An advertiser might set bids based on expected clickthrough and conversion probabilities for different ad positions (we refer to this as *bidding*

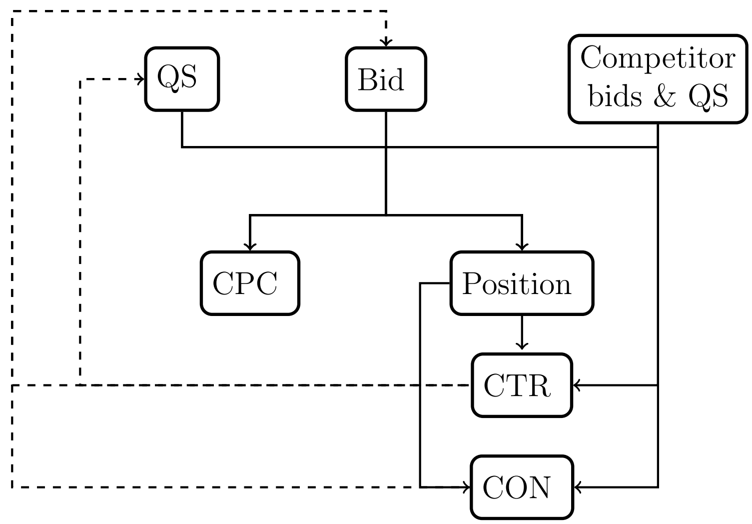


Figure 2: Conceptual model of keyword performance. QS denotes quality score, CPC denotes cost-per-click, CTR denotes clickthrough rate, and CON denotes conversion rate. Solid lines represent contemporary causal effects, dashed lines represent future causal effects.

endogeneity).

To correct for all these sources of endogeneity, the earlier mentioned studies use parametric simultaneous equations models of the clickthrough and conversion probabilities and the ad's position plus a specific strategy to solve for bidding endogeneity. Agarwal et al. (2011) use data on randomized bids to explain the position. Alternatively, Rutz et al. (2012) use latent instrumental variables (LIVs) to explain the position. In the LIV approach, the endogenous variable (ad position) is split into a part that is uncorrelated with the error terms of the clickthrough and conversion equations, the latent instruments, and a part that is potentially correlated. Finally, Ghose and Yang (2009) simultaneously model the firm's bid with the clickthrough and conversion probabilities and the ad's position.

In this paper we opt for the approach by Ghose and Yang (2009). Although randomized bids as used by Agarwal et al. (2011) yield a better source of variation to identify the causal impact of position, it is rare to find firms who actually practice randomized bidding. A drawback of the LIV approach of Rutz et al. (2012) is that it relies on the existence of latent "groups" that are correlated with position and uncorrelated with the unexplained parts of the clickthrough and conversion probabilities. In general it is unknown whether such groups exist.

The above mentioned studies find mixed results regarding the drivers of keyword performance. Generally, they agree that the more prominent the position, the higher the clickthrough probability (Ghose & Yang, 2009, Agarwal et al., 2011, and Rutz et al., 2012). Furthermore, Agarwal et al. (2011) and Ghose and Yang (2009) find that profits are usually not highest in the top positions. Instead, profits increase until some position when going down the search results page after which they decrease again. These studies ignore dynamics other than day-of-the-week effects.

From the search engine's perspective there is also literature on modeling clickthrough probabilities. These models are used to estimate quality scores of ads to help allocate ads on the search results pages. The models proposed in this literature are predictive models, no steps are taken to account for position or bidding endogeneity. One such model that allows for dynamic performance is proposed in Graepel, Candela, Borchert, and Herbrich (2010), who develop a Bayesian model for clickthrough probabilities. This model allows for dynamics by adjusting the parameters as new data comes in through a Bayesian learning algorithm that gives higher weight to more recent observations.

3. General structure of data

Google provides advertisers with a number of ad performance metrics. These metrics are aggregated on the level of the keyword and some time period, such as the hour of the day or day of the week. The performance metrics include the number of obtained impressions, clicks, and conversions, the average position over the impressions, and the average cost-per-click (CPC). Google also provides four metrics related to the quality score: quality score (ranging from 1 to 10), landing page experience, ad relevance, and expected clickthrough rate. The quality score metric is, however, not the actual measure used by Google in real-time to assign positions to ads.[3]

Based on the words in the keyword, an advertiser can construct semantic characteristics of keywords. These characteristics might be useful in estimating keyword performance. They can include the number of words in the keyword, or the specificity of the keyword (e.g. generic, branded or retailer-specific search like in Ghose & Yang, 2009). In addition, the advertiser knows the match type of each keyword. The match type of a keyword refers to how "well" the keyword must match the consumers search in order to be eligible to show, and is one of 'exact', 'broad', or 'phrase'. The broader the match type, the more divergent the search phrase and the keyword may be.

[3]For more information on the quality score, see https://support.google.com/google-ads/answer/7050591?hl=en.

4. Methods

In this section, we discuss the statistical model we propose for keyword performance. We consider model specification, parameter identification, and model inference.

The model we propose is a dynamic Bayesian model of the consumers' clickthrough and conversion behavior, the search engine's position allocating behavior, and the firm's bid behavior. The model is a time-varying parameters model, also known as a model in state-space representation (e.g. Hamilton, 1994, Chapter 13).

4.1. Model specification

We index the keywords by $i = 1, ..., N$ and time periods by $t = 1, ..., T$. We denote by I_{it}, N_{it}, and M_{it} the number of impressions, clicks, and conversions of keyword i at time t, respectively. By definition, $I_{it} \geq N_{it} \geq M_{it}$. Furthermore, we denote by p_{it}^{CTR} the unobserved clickthrough probability on keyword i at time t (the probability of a click given an impression) and by p_{it}^{CON} the conversion probability (the probability of a conversion given a click). Let POS_{it} denote the average ad position over the impressions on keyword i at time t, BID_{it} the bid, CPC_{it} the cost-per-click, and QS_{it} the quality score. Finally, let x_i denote a $(K \times 1)$ vector of (semantic) characteristics of keyword i, and s_t a vector of seasonal dummies.

We assume that, conditional on $\{p_{it}^{CTR}\}_{t=1}^{T}$ and $\{p_{it}^{CON}\}_{t=1}^{T}$, the impressions and clicks on keyword i at time t are independent across ads served, and that the impressions and clicks on keyword i are independent across time and independent of other keywords $j \neq i$. Then, the number of clicks on keyword i at time t and the number of conversions are conditionally binomially distributed. That is,

$$N_{it}|I_{it}, p_{it}^{CTR} \sim BIN(I_{it}, p_{it}^{CTR}),$$
$$M_{it}|N_{it}, p_{it}^{CON} \sim BIN(N_{it}, p_{it}^{CON}),$$

for $i = 1, ..., N$ and $t = 1, ..., T$.

Next, we propose a dynamic simultaneous equations model of the clickthrough and conversion probabilities, the ad's position, and the firm's bid. The model is given by

$$
\begin{aligned}
p_{it}^{CTR} &= \Lambda\left(\alpha_i^{CTR} + x_i'\beta_t^{CTR} + \lambda_i^{CTR}\ln(\mathrm{POS}_{it}) + s_t'\gamma^{CTR} + \eta_{it}^{CTR}\right), && (1)\\
p_{it}^{CON} &= \Lambda\left(\alpha_i^{CON} + x_i'\beta_t^{CON} + \lambda_i^{CON}\ln(\mathrm{POS}_{it}) + s_t'\gamma^{CON} + \eta_{it}^{CON}\right), && (2)\\
\ln(\mathrm{POS}_{it}) &= \alpha_i^{POS} + x_i'\beta_t^{POS} + \lambda_i^{POS}\ln(\mathrm{BID}_{it}) + \psi^{POS}\ln(\mathrm{QS}_{it}) + s_t'\gamma^{POS} + \eta_{it}^{POS}, && (3)\\
\ln(\mathrm{BID}_{it}) &= \alpha_i^{BID} + x_i'\beta_t^{BID} + q_{it}'\delta_i^{BID} + s_t'\gamma^{BID} + \eta_{it}^{BID}, && (4)
\end{aligned}
$$

for keywords $i = 1, ..., N$ and time periods $t = 1, ..., T$, where $\Lambda(\theta) \equiv 1/(1+\exp(-\theta))$ is the standard logistic link function.

The key parts of the model are the clickthrough and conversion equations (1) and (2). These two equations have an equivalent functional form with different parameters. The logistic link function is used to map real-valued numbers to probabilities between 0 and 1. The baseline levels of the clickthrough and conversion probabilities are captured in the keyword-specific intercepts α_i, discussed in detail in Section 4.1.2. Stochastic dynamics are captured in the term $x_i'\beta_t$, which captures the time variation in clickthrough and conversion probabilities for different types of keywords. The process for β_t captures the carryover effects of shocks to subsequent periods, and is discussed in Section 4.1.1. Deterministic seasonal effects are captured in the term $s_t'\gamma$. The effect of ad position is captured in the keyword-specific parameters λ_i and is discussed in Section 4.1.2.

The position equation (3) is included to correct for position endogeneity and the bid equation (4) to correct for bidding endogeneity. For both equations we use a linear specification for the log transformed variables. They also deviate from the clickthrough and conversion equations in that we let the position depend on the bid and quality score, and let the bid depend on q_{it}, a vector of instrumental variables not included in the other equations. We discuss these instruments in Section 4.1.3. Note that the position equation (3) can be rewritten as

$$
\mathrm{POS}_{it} = g_{it}^{POS}\mathrm{QS}_{it}^{\psi^{POS}}\mathrm{BID}_{it}^{\lambda_i^{POS}},
$$

where the multiplication factor g_{it}^{POS} depends on x_i, s_t and η_{it} in a potentially time-varying way. Hence, we assume that the position depends on the bid and the quality score in a multiplicative way. The parameters ψ^{POS} and λ^{POS} are elasticity parameters; if the bid increases by 1%, then the position increases by λ^{POS}%.

The key elements in correcting for position endogeneity are the keyword- and time-specific error terms in $\eta_{it} = (\eta_{it}^{CTR}, \eta_{it}^{CON}, \eta_{it}^{POS}, \eta_{it}^{BID})'$. We assume that η_{it} is multivariate normally distributed for keyword i and time t and independent across keywords and time. That is,

$$\eta_{it} \sim MVN(0, \Sigma_\eta), \tag{5}$$

where all elements of the positive definite matrix Σ_η are allowed to be non-zero.

Even when no position endogeneity is present, it is important to include η_{it} into the clickthrough and conversion equations. The model is based on the aggregation of choices on the keyword- and time-level. The parameter η_{it} captures keyword- and time-specific deviations that are not captured by other model parameters. In case a keyword i receives many observations and clicks in a given time period t, the likelihood of p_{it}^{CTR} and p_{it}^{CON} given the observed data is highly peaked at the observed fractions of clicks/conversions. Hence, the estimation procedure will model the clickthrough and conversion probabilities to be (almost) equal to the realized proportions in the data. In case η_{it}^{CTR} and η_{it}^{CON} are included, they can capture potential deviations between expected and realized proportions. In case they are not included, the estimates of the parameters will become such that they mainly fit these few observations with many impressions, instead of representing general patterns across the whole set of keywords.

4.1.1. Time-varying parameters: the dynamic impact of shocks

The impact of changes in the environment on ad performance is captured in the time-varying parameters $\beta_t = (\beta_t^{CTR'}, \beta_t^{CON'}, \beta_t^{POS'}, \beta_t^{BID'})'$. Changes in the environment can result from changes in macroeconomic conditions, in the firm (e.g. changing reputation), in the market competitiveness (e.g. new competitor or the launch of a new product), in the search engine's position-allocating mechanism, or in consumers (e.g. changing tastes and attitudes). The effect of changes on ad performance can be transitory or permanent.

To capture the dynamics in SEA environments, we take independent AR(1) processes for the time-varying parameters. That is,

$$\beta_{t+1} = \Phi\beta_t + \nu_t, \qquad \nu_t \sim MVN(0, \Sigma_\beta), \qquad \beta_1 \sim MVN(0, 5\Sigma_\beta), \tag{6}$$

for $t = 1, ..., T$, where Φ and Σ_β are $(4K \times 4K)$ diagonal matrices. These AR(1) processes can capture a wide variety of paths for the time-varying parameters (Van Heerde, Mela, & Manchanda, 2004).

The autoregressive parameters $\{\phi_k\}_{k=1}^{4K}$ on the diagonal of Φ measure the persistence of the impact of shocks on future values of β_{kt}. In case $\phi_k = 1$, shocks are permanent. In case

$\phi_k = 0$, shocks do not impact future clickthrough and conversion probabilities, and a static model would do. In case $0 < \phi_k < 1$, the effects of shocks carry over to next periods but the process is mean-reverting: shocks die out geometrically with decay rate ϕ_k.

The AR(1) processes in Equation (6) have no intercept. If the β_t series is stationary, an intercept captures the unconditional mean of the series. If the series is nonstationary, the intercept would either capture the level at time $t = 1$ (in deviation-from-mean form), or a drift parameter (in regular form). We move this intercept to the mean of the α_i parameter as will be explained next. This ensures that we can interpret the intercept as in the deviations-from-mean form and thus have equal interpretation of the parameter in case of stationarity and nonstationarity of the β_t series. Moreover, it helps improve the mixing rates of the sampler.

4.1.2. Unobserved heterogeneity across keywords

The model captures unobserved heterogeneity across keywords through the keyword-specific parameters $\alpha_i = (\alpha_i^{CTR}, \alpha_i^{CON}, \alpha_i^{POS}, \alpha_i^{BID})'$ and $\lambda_i = (\lambda_i^{CTR}, \lambda_i^{CON}, \lambda_i^{POS})'$. We shrink α_i and λ_i to common means across similar keywords.

The parameters in α_i capture common baseline levels of ad performance as well as keyword-specific deviations. We take α_i to be independently normally distributed across keywords,

$$\begin{pmatrix} \alpha_i^{CTR} \\ \alpha_i^{CON} \\ \alpha_i^{POS} \\ \alpha_i^{BID} \end{pmatrix} \sim MVN \left(\begin{pmatrix} x_i'\tilde{\alpha}^{CTR} \\ x_i'\tilde{\alpha}^{CON} \\ x_i'\tilde{\alpha}^{POS} \\ x_i'\tilde{\alpha}^{BID} \end{pmatrix}, \begin{bmatrix} \sigma^2_{\alpha,CTR} & 0 & 0 & 0 \\ 0 & \sigma^2_{\alpha,CON} & 0 & 0 \\ 0 & 0 & \sigma^2_{\alpha,POS} & 0 \\ 0 & 0 & 0 & \sigma^2_{\alpha,BID} \end{bmatrix} \right), \tag{7}$$

for $i = 1, ..., N$, where $\tilde{\alpha} = (\tilde{\alpha}^{CTR}, \tilde{\alpha}^{CON}, \tilde{\alpha}^{POS}, \tilde{\alpha}^{BID})$ captures the common baseline levels.

The parameters in λ_i capture the impact of ad position on clickthrough and conversion probabilities (λ_i^{CTR}, λ_i^{CON}), and the effect of bid on ad position (λ_i^{POS}). We take λ_i to be independently normally distributed across keywords. That is,

$$\begin{pmatrix} \lambda_i^{CTR} \\ \lambda_i^{CON} \\ \lambda_i^{POS} \end{pmatrix} \sim MVN \left(\begin{pmatrix} x_i'\tilde{\lambda}^{CTR} \\ x_i'\tilde{\lambda}^{CON} \\ x_i'\tilde{\lambda}^{POS} \end{pmatrix}, \begin{bmatrix} \sigma^2_{\lambda,CTR} & 0 & 0 \\ 0 & \sigma^2_{\lambda,CON} & 0 \\ 0 & 0 & \sigma^2_{\lambda,POS} \end{bmatrix} \right), \tag{8}$$

for $i = 1, ..., N$.

4.1.3. Instrumental variables

The instruments q_{it} in the bid equation (4) are necessary for identification of the dynamic simultaneous equations model. They must be excluded from the other equations (1)-(3). A researcher can take any set of valid instruments: the instruments should be correlated with the bid, but uncorrelated with ad position and clickthrough and conversion probabilities after correcting for bid/position.

We propose to use previous performance indicators as instruments, as these indicators capture the potential strategic bidding behavior of advertisers that causes the bidding endogeneity. We consider the previous clickthrough rate and the previous number of impressions obtained.[4] For the previous clickthrough rate to be a valid instrument, we assume that the quality score measure we include in the position equation is a sufficient statistic for the previous clickthrough rate in explaining ad position.

To allow for heterogeneity in the effect of the instruments we consider keyword-specific parameters

$$\delta_i^{BID} \sim MVN(x_i'\tilde{\delta}^{BID}, \Sigma_\delta^{BID}), \tag{9}$$

where Σ_δ^{BID} is a positive definite diagonal matrix.

4.2. Parameter identification

To ensure that the parameters in the model in (1)-(9) are identified, we have to consider two issues. First, for the stochastic dynamics part $x_i'\beta_t$, a researcher may wish to include many characteristics such that the matrix $X = (x_1, x_2, x_3, ..., x_N)'$ is not of full column rank. For example, a researcher may want to include an intercept and all dummies for a categorical variable, to distinguish between market-level shocks and the lower level shocks for different categories. In this case, where X is not of full column rank, identification restrictions need to be imposed. More specifically, a set of variables k^* has to be selected, such that the matrix X without the columns corresponding to these variables in k^* is of full column rank. For these variables in k^*, the following restrictions are sufficient for identification: (i) $\beta_1 = 0$, (ii) $\tilde{\alpha} = 0$, and (iii) $\tilde{\lambda} = 0$. Note that these variables will still have a non-zero effect for $t > 1$.

Second, the simultaneous equations model in (1)-(4) is identified as the model is a tri-

[4] Ghose and Yang (2009) use the lagged ad position as instrument in the bid equation. Exogeneity of this instrument depends on the assumption that the error terms in the position equation are serially uncorrelated. This assumption might very well be unrealistic, rendering lagged ad position invalid as instrument.

angular system (Greene, 2012, Ghose & Yang, 2009): the bid equation depends only on exogenous variables, the position equation depends only on the endogenous variable bid, and the clickthrough and conversion equations depend only on the endogenous variable ad position. Identification in this triangular system is ensured through the exclusion restrictions that the instrumental variables in the bid equation are excluded in the clickthrough, conversion, and position equations, and the bid variable in the position equation is excluded in the clickthrough and conversion equations. Hence, the model is identified and we do not need to impose restrictions on the covariance matrix Σ_η.

4.3. Bayesian inference

We perform Bayesian inference for the dynamic simultaneous equations model in (1)-(9). We use Markov Chain Monte Carlo (MCMC) techniques and rely on a Gibbs sampler with Polya-Gamma data augmentation (Geman & Geman, 1987, Tanner & Wong, 1987, Polson et al., 2013). The advantage of using a Bayesian estimation approach is that we can use informative priors for keyword characteristics that are very rare. The Gibbs sampler also deals naturally with missing values.

The Polya-Gamma data augmentation scheme is suitable for binomial logistic regression models (Polson et al., 2013). It allows for exact inference by introducing one layer of Polya-Gamma distributed latent variables, where the latent variables are drawn at the level of the keyword and time period. Alternative MCMC approaches for Bayesian inference for logistic regression models are (i) data augmentation schemes where the logistic distributed error terms are approximated by mixtures of normals (Holmes & Held, 2006, Frühwirth-Schnatter & Frühwirth, 2010) or (ii) an independence or random walk Metropolis-Hastings (MH) algorithm without data augmentation (Rossi, Allenby, & McCulloch, 2005). The disadvantages of the alternative data augmentation schemes are that they are not exact, require two layers of auxiliary variables, and require much more memory storage as augmentation is performed on the level of an impression or click and not on the total number of impressions and clicks (this is especially relevant for the SEA application). The disadvantage of the MH algorithms is that they often have poor mixing rates and that tuning may be required (Frühwirth-Schnatter & Frühwirth, 2010). This is especially important when dynamic states are involved.

The Gibbs sampler we use is outlined in Appendix A. In this Gibbs sampler, we subsequently draw the auxiliary variables from the Polya-Gamma distribution, the time-varying parameters from a multivariate normal distribution using the forward-filtering backward-sampling (FFBS) algorithm (Carter & Kohn, 1994, Frühwirth-Schnatter, 1994) of Durbin and Koopman (2002) and collapsed filtering (Durbin & Koopman, 2012, Jungbacker & Koopman, 2015), the time-invariant parameters from multivariate normal distributions, and the covariance matrices from inverse Wishart distributions. Specialized code is written in R (R Core Team, 2013) and C++ to enhance computation speed using the R package Rcpp (Eddelbuettel & François, 2011).

5. Empirical application

In this section, we apply the proposed dynamic Bayesian model to data of a Dutch online retailer. We present the data in Section 5.1 and discuss the in-sample results in Section 5.2. Finally, in Section 5.3, we compare the performance of our dynamic model to a static model without time-varying parameters.

5.1. Data

The data contains the historical performance of 14,710 keywords related to laptops measured at the daily frequency over the period January 1, 2014 until March 31, 2016.[5] The data contains information on the daily number of impressions, clicks, and conversions[6], and the daily average cost-per-click, ad position, and quality score[7]. We consider all data for model inference. In total, the keywords obtained 47.0 million impressions, 1.6 million clicks and 33.0 thousand conversions. The average clickthrough rate was 3.4% and the average conversion rate was 2.0%. Moreover, the top 5% of keywords based on impressions accounted for 92.5% of total impressions, whereas the bottom 50% accounted for 0.2% of total impressions.

We also use semantic characteristics of keywords. Each keyword is assigned to one of four categories indicating the specificity of the keyword: (i) 'generic', (ii) 'brand only', (iii) 'brand and series', or (iv) 'retailer'. The 'brand only' keywords are keywords that include the brand name of a laptop but not the name of a specific series or model (e.g. 'asus laptop'),

[5] Google provides data aggregated on the device used by the consumer (computer, tablet, or mobile device). We only include data on searches made via the computer, as consumer behavior might differ for the three electronic devices and the comparative usage of the three devices might have changed over time.

[6] Conversions are measured based on the keyword associated with the *last clicked ad* by the consumer as tracked by Google. Conversions are counted when the consumers makes a purchase within 30 days of clicking on the last clicked ad.

[7] We do not have data on the landing page experience, ad relevance, and expected clickthrough rate. Furthermore, we impute missing quality scores with a 6. Quality scores are missing when there were insufficient previous impressions and clicks for Google to determine the quality score. In these cases, Google uses a quality score of 6 in the keyword auction, see https://searchengineland.com/google-adwords-keyword-quality-score-reporting-update-226355.

whereas the 'brand and series' keywords include at least a brand's series name (e.g. 'asus vivobook'). We divide the keywords in the 'brand only' and 'brand and series' categories into the eight brands available at the retailer: Acer, Apple, Asus, HP, Lenovo, Microsoft, MSI, and Toshiba.

Furthermore, we know the match type of the keyword, which is either broad or exact, and the number of words in the keyword. For these two variables we consider time-invariant parameters, that is, $\beta_t = 0$. We include seasonal dummies for the day of the week.

As instruments, we use the previous clickthrough rate and the natural logarithm of the previously obtained number of impressions. We compute the previous clickthrough rate by taking the realized clickthrough rate over the previous month.[8] Once the advertiser has implemented the model to set the bid, the previous clickthrough rate can be estimated using Equation (1) instead. Furthermore, for the previous impressions we consider the average daily number of impressions obtained on a keyword in the previous month.

Finally, we use the CPC as a proxy for the bid as done in Ghose and Yang (2009) and Skiera and Abou Nabout (2013). Data on historical bids are not provided by Google, and have not been stored by the retailer. Using the CPC instead of the bid is justified for competitive keywords, where the difference between the CPC and the bid is small (Abou Nabout, Skiera, Stepanchuk, & Gerstmeier, 2012). A disadvantage is that we do not always observe the CPC when we observe the position. We therefore impute the missing CPCs for explaining position, using a stochastic local level model.[9]

5.2. Baseline results

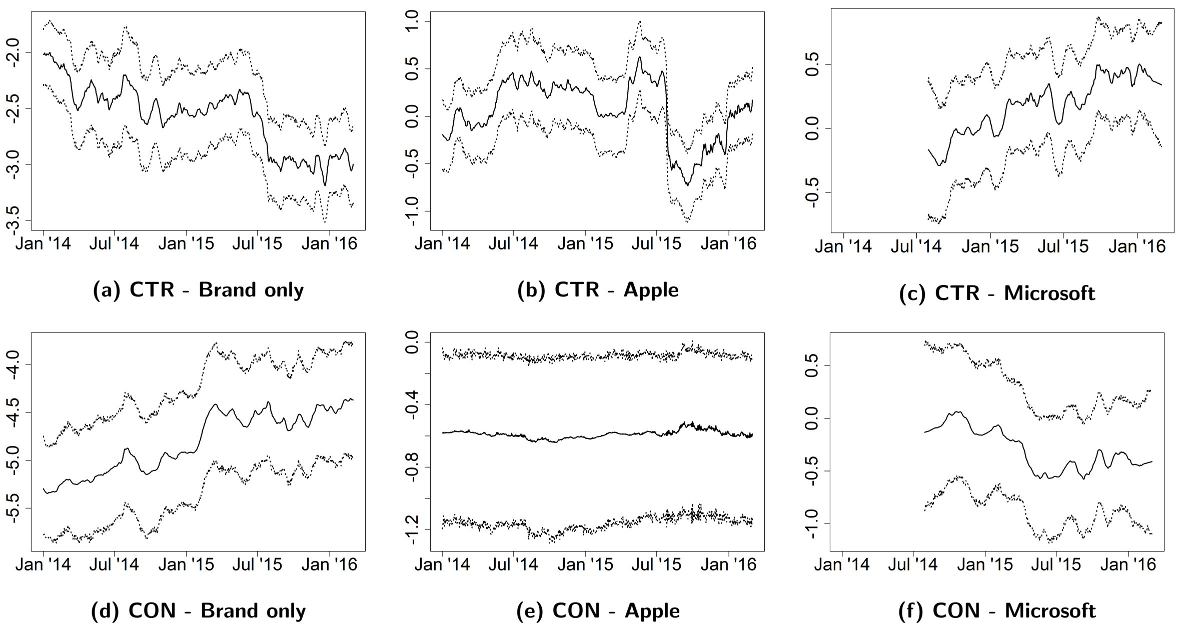
In this section, we discuss the in-sample results for the proposed dynamic Bayesian model. Posterior results are obtained using 35.000 simulations after 5.000 burn-in draws. We keep every 4^{th} draw to deal with the correlation in the chain. Here, we show the most important results.

We find that clickthrough and conversion probabilities have substantially changed over time. Figure 3 shows illustrative examples of the smoothed estimates and 95% pointwise highest posterior density intervals (HPDIs) of the time-varying parameters. For the brand

[8]In case a keyword obtained at least 5,000 impressions in the previous month, we take the clickthrough rate (CTR) of that *specific keyword*. Otherwise, we take the CTR over the *campaign group* the keyword was assigned to or, if that campaign group received less than 5,000 impressions in the previous month, the *specificity category* the keyword was assigned to.

[9]The local level model is given by $CPC_{i,t+1} = \mu_{it}$ with $\mu_{i,t+1} = \mu_{it} + \varepsilon_{it}$, $\mu_{i1} \sim N(\overline{CPC}, 0.5)$, and $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$, where $\overline{CPC}$ is the mean of all CPCs in the dataset and we estimate σ_ε^2 with maximum likelihood ($\hat{\sigma}_\varepsilon^2 = 0.007$ based on a set of popular keywords). We use the smoothed estimates of CPC_{it}.

only keywords, we find that clickthrough probabilities have substantially decreased over time, whereas conversion probabilities have increased. The conversion performance of Apple keywords was stable, whereas the clickthrough performance was volatile. For Microsoft laptops, introduced at the retailer in August 2014, we see quite some time variation in ad performance with alternating periods of high and low clickthrough and conversion probabilities.



(a) CTR - Brand only

(b) CTR - Apple

(c) CTR - Microsoft

(d) CON - Brand only

(e) CON - Apple

(f) CON - Microsoft

Figure 3: Posterior means and 95% highest posterior density intervals of $\{\beta_t\}_{t=1}^T + \tilde{\alpha}$ for the brand only keywords, retailer-specific keywords, and Microsoft keywords. For both the clickthrough (CTR) and conversion (CON) probabilities.

The 95% HPDIs in Figure 3 are quite wide. This is not so much caused by uncertainty in the dynamics in the time-varying parameter series (as the different smoothed draws follow similar dynamics), but is mainly caused by uncertainty in how the absolute levels should be attributed to the higher-level brand only effect and the lower-level brand effects (e.g. Apple, Microsoft). Adding the brand only effect to any of the brand effects, we find much smaller 95% HPDIs.

To assess the persistence of shocks — how long shocks carryover to next periods — we consider the posterior results for the autoregressive parameters in Φ in Table 1. We also compute the half-life of shocks. The half-life is the number of weeks before the effect of the shock is below 50% from its original value.[10] For both the clickthrough and conversion probabilities, we find that shocks at the specificity level are permanent or highly persistent.

[10]The half-life of shocks (in days) d can be computed by equating $\phi^d = 0.5$, that is, $d = \ln(0.5)/\ln(\phi)$ where ϕ is the posterior mean of the autoregressive parameter.

	CTR				CON			
		2.5th	*97.5th*	*Half-life*		*2.5th*	*97.5th*	*Half-life*
	Mean	*percentile*	*percentile*	*(in weeks)*	*Mean*	*percentile*	*percentile*	*(in weeks)*
Generic	1.00	0.99	1.00	-	1.00	0.98	1.00	-
Brand only	1.00	0.99	1.00	-	1.00	0.98	1.00	-
Brand & series	1.00	0.99	1.01	-	1.00	0.99	1.00	-
Retailer	1.00	0.99	1.00	-	0.99	0.96	1.01	10.1
Microsoft	0.99	0.97	1.00	11.6	0.98	0.94	1.01	4.6
Toshiba	0.99	0.97	1.00	7.2	0.30	-0.83	1.00	0.1
HP	1.00	0.99	1.00	-	0.75	-0.54	1.01	0.3
Acer	0.97	0.92	1.00	3.2	0.52	-0.73	1.01	0.2
Asus	0.68	-0.11	1.01	0.3	0.43	-0.74	1.00	0.1
Apple	0.99	0.98	1.00	11.5	0.30	-0.82	1.00	0.1
Lenovo	0.99	0.98	1.00	14.8	0.35	-0.70	0.99	0.1
MSI	0.99	0.96	1.00	6.9	0.35	-0.84	1.00	0.1

Table 1: Posterior means, 95% highest posterior density intervals, and the half-life of shocks (based on the posterior means) for the autoregressive parameters Φ in the AR(1) process for $\{\beta_t\}_{t=1}^T$ in Equation (6) for the clickthrough (CTR) and conversion (CON) probabilities.

The brand-level shocks on clickthrough performance are generally also persistent, with permanent shocks for HP ads and a half-life ranging from 0.3 weeks for Asus ads to 14.8 weeks for Lenovo ads. The effects of brand-level shocks on the conversion performance are more transitory; the half-life ranges from 0.1 weeks to 4.6 weeks.

To compare the relative performance of keywords, in Figure 4 we plot the time-varying parameter series including baseline levels for the different specificity groups (left) and brands (right). Clickthrough probabilities are highest for retailer-specific keywords, followed by generic, brand only, and brand & series keywords. Conversion probabilities are also highest for retailer-specific keywords followed by brand & series, generic and brand only keywords. The clickthrough probabilities of different brands are volatile, whereas the conversion probabilities are relatively stable. Conversion probabilities are lowest for Apple keywords, followed by Microsoft, MSI, and Toshiba keywords. All graphs again show quite some time variation.

Table 2 displays day-of-the-week effects and the effects of keyword length and match type. Clickthrough probabilities are highest on Mondays to Wednesdays and for shorter and exact keywords. Conversion probabilities are lowest on Saturdays, and highest for longer and exact keywords. The posterior estimates for the standard deviations σ_α show that there is substantial variation across keywords in the baseline level of clickthrough and conversion

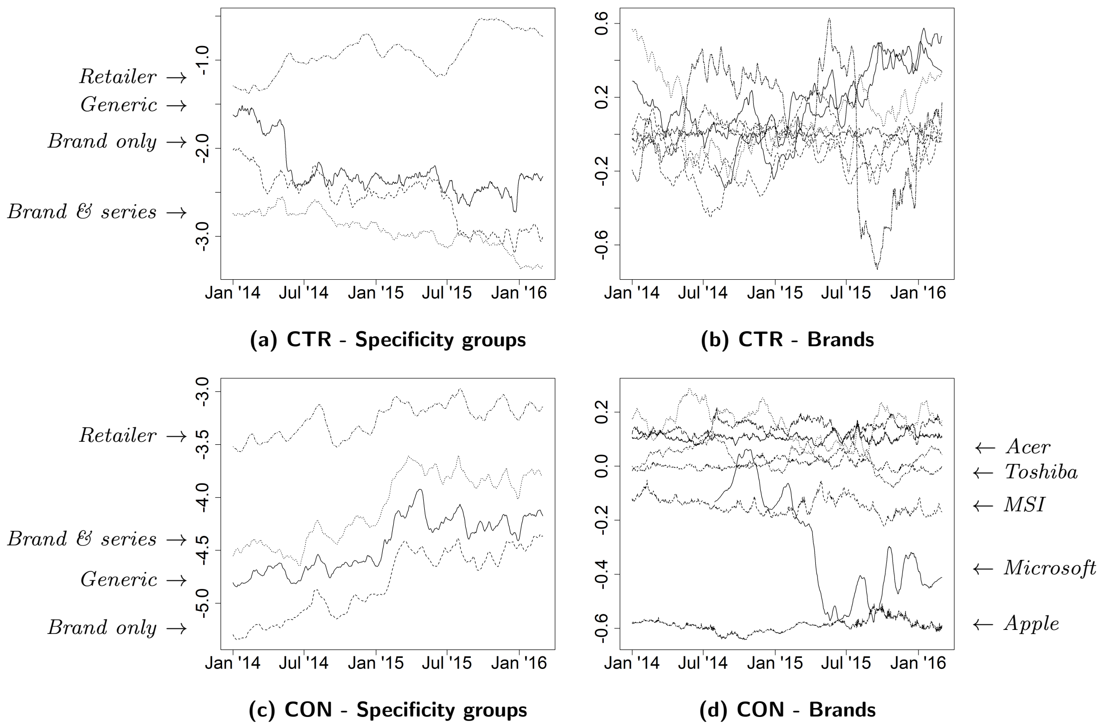


Figure 4: Plots of the series $\{\beta_t\}_{t=1}^{T} + \tilde{\alpha}$ for the specificity and brand series, for the clickthrough (CTR) and conversion (CON) probabilities.

probabilities.

Next, we consider the effect of ad position. Table 3 displays the estimated effect of ad position (columns 2 and 3) and of bid/CPC (column 4). We find that the more prominent the ad — the lower the position number — the higher the clickthrough probability. This holds in general for all types of keywords, although the relationship is strongest for retailer-specific and exact keywords, and weakest for Apple, HP, and MSI keywords. For the conversion probabilities we do not find strong evidence that ad position affects conversion probabilities in general. Furthermore, high bids/CPCs are mostly associated with more prominent ads. This holds strongest for long and Microsoft keywords. The estimates for the standard deviations σ_λ show that the effect of position on clickthrough and conversion probabilities varies substantially over similar keywords.

Table 3, columns 5 and 6, show the posterior results for the instruments in the bid equation. The results indicate that the bid/CPC is associated with past performance. The instruments seem strong enough to identify the other parameters. In general, the advertiser sets higher bids on keywords that previously obtained a high number of impressions and a high clickthrough rate. Given the size of the variance across keywords (σ_δ), the reverse relationship also seems to hold for a number of keywords. This implies that the advertiser may

	CTR	CON
Tuesday	0.00 (0.01)	0.01 (0.03)
Wednesday	-0.01 (0.01)	-0.03 (0.03)
Thursday	-0.02 (0.01)	-0.03 (0.03)
Friday	-0.03 (0.01)	-0.01 (0.04)
Saturday	-0.06 (0.01)	-0.05 (0.03)
Sunday	-0.06 (0.01)	-0.03 (0.03)
ln (# words)	-0.04 (0.02)	0.16 (0.04)
Exact match	0.48 (0.02)	0.25 (0.04)
σ_α	0.65 (0.01)	0.43 (0.02)

Table 2: Posterior means and standard deviations (in parentheses) for the seasonal effects (γ), the time-invariant parameters in $\tilde{\alpha}$ and σ_α (the square root of the diagonal of Σ_α).

	Impact POS on CTR	Impact POS on CON	Impact BID/CPC on POS	Impact lagged CTR on BID	Impact lagged IMP on BID
Intercept	-1.01 (0.04)	0.07 (0.07)	-0.04 (0.02)	0.02 (0.00)	0.07 (0.01)
Brand only	-0.15 (0.05)	0.06 (0.11)	0.05 (0.02)	-0.01 (0.01)	0.03 (0.02)
Brand & series	0.03 (0.04)	-0.05 (0.10)	0.15 (0.02)	-0.01 (0.01)	-0.03 (0.01)
Retailer	-0.46 (0.11)	-0.05 (0.15)	0.22 (0.05)	-0.01 (0.01)	-0.22 (0.03)
Microsoft	0.08 (0.08)	0.29 (0.15)	-0.14 (0.03)	-0.02 (0.01)	0.00 (0.02)
Toshiba	-0.02 (0.04)	0.02 (0.12)	-0.08 (0.02)	-0.02 (0.01)	0.04 (0.01)
HP	0.12 (0.04)	-0.04 (0.10)	-0.01 (0.01)	0.00 (0.01)	0.04 (0.01)
Asus	0.09 (0.04)	-0.16 (0.10)	-0.01 (0.01)	0.02 (0.01)	0.01 (0.01)
Apple	0.21 (0.05)	0.07 (0.13)	-0.04 (0.02)	-0.01 (0.01)	-0.03 (0.02)
Lenovo	0.00 (0.04)	-0.30 (0.11)	-0.02 (0.02)	0.03 (0.01)	0.05 (0.01)
MSI	0.14 (0.06)	0.17 (0.17)	0.02 (0.02)	-0.03 (0.01)	0.01 (0.01)
ln(# words)	-0.02 (0.03)	0.03 (0.06)	-0.07 (0.01)	0.00 (0.00)	0.00 (0.01)
Exact match	-0.22 (0.02)	0.07 (0.05)	-0.05 (0.01)	-0.01 (0.00)	-0.01 (0.01)
$\sigma_\lambda/\sigma_\delta$	0.43 (0.01)	0.18 (0.02)	0.28 (0.00)	0.05 (0.00)	0.13 (0.00)

Table 3: Posterior means and standard deviations (in parentheses) for the position parameters in the clickthrough and conversion equations (columns 2-3), the bid parameters in the position equation ($\tilde{\lambda}$) (column 4), the instruments in the bid equation ($\tilde{\delta}$) (columns 5-6) and σ_λ and σ_δ (the square root of diagonal of Σ_λ and Σ_δ respectively). Base categories are specificity 'Generic', brand 'Acer' and match type 'Broad'.

not always bid strategically based on previous clickthrough rates and impressions obtained.

Table 4 shows the posterior mean of the covariance matrix of the error terms, Σ_η: the variances are displayed on the diagonal, the covariances on the upper diagonal, and the correlations on the lower diagonal. Position endogeneity seems present, as the unexplained parts of the clickthrough probabilities are positively correlated with the unexplained parts of ad position. For conversion probabilities we find no strong evidence of position endogeneity.

	CTR	CON	POS	BID/CPC
CTR	**0.137**	**-0.003**	**0.027**	**0.003**
	0.002	0.004	0.002	0.001
CON	**-0.045**	**0.026**	**0.005**	**-0.003**
	0.067	0.005	0.005	0.003
POS	**0.222**	**0.099**	**0.106**	**-0.030**
	0.016	0.103	0.000	0.001
BID	**0.018**	**-0.048**	**-0.240**	**0.144**
	0.008	0.049	0.010	0.001

Table 4: Posterior means and standard deviations (in parentheses) for the variances (diagonal), covariances (upper diagonal) and correlations (lower diagonal) of η_{it}

.

Finally, bidding endogeneity also seems present, as there is a negative correlation of -0.240 between the position and bid error terms. Part of this correlation can be explained because we use the CPC to proxy the bid; the ad position and CPC are both influenced by unobserved competitive behavior. These findings reinforce that it is important to account for these forms of endogeneity.

5.3. Model comparison

In this section, we compare the performance of the dynamic model to a static model with seasonality, that is, setting all $\beta_t = 0$. We compute log Bayes factors to evaluate the models' in-sample and out-of-sample performance. In case the log Bayes factor is greater than log(3) we have sufficient evidence to favor the null model (the static model), in case it is smaller than log(1/3) we have evidence to favor the alternative model (the dynamic model) (Kass & Raftery, 1995).

We compute the in-sample log Bayes factors with the Savage-Dickey density ratio, using the estimates obtained from the dynamic model only (Dickey, 1971).[11] We compute the

[11] The in-sample log Bayes factor of the static model against the dynamic model can be computed by the

predictive Bayes factors using predictions from both the dynamic and the static model for a full year. For these predictions we use a moving window of 26 weeks; after each window we make predictions for each day in the next week and then we move the window one week further. This also allows the parameters of the static model to change. We make predictions for each day in the period March 30, 2015 until March 27, 2016. For each model, we use 8.000 simulations after 2.000 burn-in draws and we keep each 4^{th} draw.

	CTR	CON
In-sample log Bayes factor	-474 364	-39 710
Predictive log Bayes factor	-6 790	-87

Table 5: Log Bayes factors for the dynamic model (alternative model) against a static model with $\beta_t = 0$ (null model).

Table 5 shows the log Bayes factors separately for the clickthrough and conversion equation. The log Bayes factors are all highly negative and much smaller than $\log(1/3)$ $(= -0.48)$. Hence, the dynamic model is superior to a static model in terms of both in- and out-of-sample performance. There is thus substantial evidence of dynamics in the clickthrough and conversion probabilities in the dataset, indicating that a dynamic SEA strategy is to be preferred over a static strategy.

Illustrative examples of the dynamic and static models' predictions are given in Figure 5. The clickthrough predictions are given in the top three figures, the conversion predictions in the bottom three figures. Overall, the predictions of the dynamic model (solid lines) are more volatile than those of the static model (dashed lines). For retailer-specific keywords we find that the dynamic model's clickthrough forecasts fluctuate around the static model's forecasts. Hence, where the dynamic model is able to capture short-term fluctuations, the static model with a moving window is not. For the conversion predictions we find that, in the period April 2015 until July 2015, the dynamic predictions are substantially higher than

Savage-Dickey density ratio

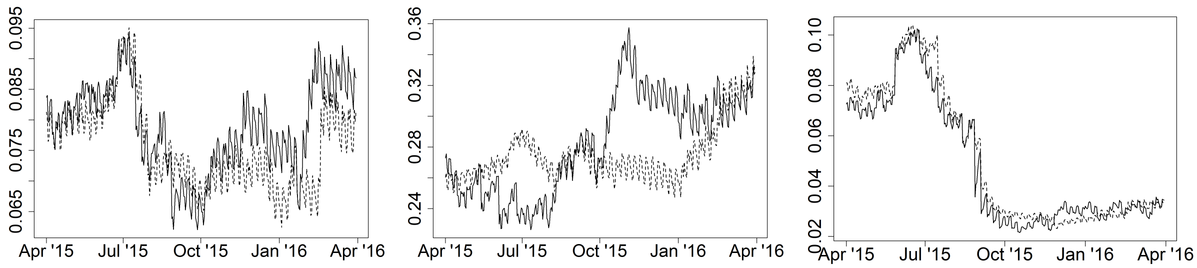
$$\ln BF = \ln p(\Phi|y)|_{\Phi=O} - \ln p(\Phi)|_{\Phi=O}, \tag{10}$$

where $p(\Phi|y)$ denotes the posterior marginal pdf of Φ from the dynamic model, and $p(\Phi)$ denotes the prior pdf of Φ. We approximate the first term on the right hand side in Equation (10) using Rao-Blackwellization based on the full conditional posterior distribution of Φ (Gelfand & Smith, 1990). That is,

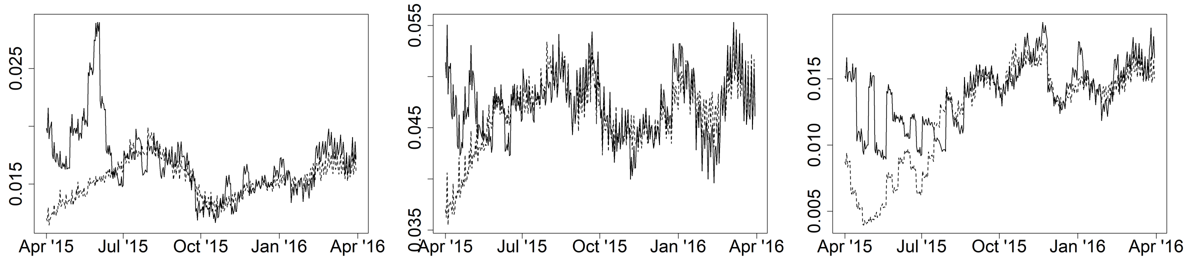
$$\ln p(\Phi|y)|_{\Phi=O} \approx \ln \left(\frac{1}{S} \sum_{s=1}^{S} p(\Phi|\beta^{(s)}, \Sigma_\beta^{(s)})|_{\Phi=O} \right), \tag{11}$$

where S is the number of Monte Carlo simulations, and $\beta^{(s)}$ and $\Sigma_\beta^{(s)}$ are the parameter draws at the s^{th} simulation.

the static predictions. The reason is reflected in Figure 4, where we see a substantial increase in the conversion rates of all specificity groups in the period October 2014 until April 2015. The dynamic model timely captures this increase whereas the static model lags behind.



(a) CTR - Generic, broad, 1 word **(b) CTR - Retailer, broad, 1 word** **(c) CTR - Apple, brand & series, exact, 3 words**



(d) CON - Generic, broad, 1 word **(e) CON - Retailer, broad, 1 word** **(f) CON - Apple, brand & series, exact, 3 words**

Figure 5: Posterior mean of predictions for clickthrough (a)-(c) and conversion (d)-(f) probabilities for three keywords for dynamic model (solid lines) and static model (dashed lines) using a moving window of 26 weeks. Ad position is set to 1.

6. Managerial implications

The managerial implications of this paper are threefold. First, advertisers can use the model to obtain accurate daily estimates of clickthrough and conversion probabilities of individual ads. These estimates can be used to set bids and test the performance of text ads and landing pages. These estimates can also be used to identify keywords of which the performance is divergent from similar keywords.

Second, advertisers can use the model to examine the extent of dynamics in their SEA environment. The more dynamic the environment and the higher the persistence of shocks, the more often the SEA strategy should be revised. Moreover, advertisers that manage large ad portfolios can prioritize their focus on keywords based on the expected influence and persistence of shocks on the keywords' performance.

Finally, advertisers can use the model to track the performance of ads to timely identify when the performance of keywords changes. An advertiser can then analyze the causes of these changes and adjust, for example, the text ad, landing page, product pricing, or bid accordingly.

As a final remark, the model's predictions of clickthrough and conversion probabilities are insufficient to determine the optimal bid per keyword. To set the optimal bid, an advertiser has to know the value of each obtained impression, click and conversion, using additional information on spillover effects to future searches (see Rutz & Bucklin, 2011, Rutz, Trusov, & Bucklin, 2011, Agarwal et al., 2011), substitution effects across marketing channels (see Yang & Ghose, 2010, Dinner, Van Heerde, & Neslin, 2014, and Blake, Nosko, & Tadelis, 2015), and branding profits of keywords (see Ghose & Yang, 2008). Combining the information from these sources requires the formation of an attribution strategy like in Li and Kannan (2014). An alternative is to use a bidding heuristic as given in Skiera and Abou Nabout (2013). We therefore consider the determination of the optimal bid to be outside the scope of this paper.

7. Summary and conclusions

In this article, we propose a dynamic Bayesian model for clickthrough and conversion probabilities of paid search advertisements. Clickthrough and conversion probabilities can be subject to changes over time, due to, for example, changes in the tastes and attitudes of consumers or the launch of a new product. Gaining insight into the dynamics of ad performance is crucial for advertisers to develop effective search engine advertising strategies.

Our main contribution is the development of a model that is especially suited to deal with dynamic SEA environments: the model allows for time-varying parameters, seasonal effects, data sparsity, missing data, position endogeneity and unobserved cross-sectional heterogeneity. Moreover, we propose AR(1) processes for the time-varying parameters, thereby allowing for shocks on different types of ads (e.g. brand-specific versus generic ads) to have different dynamic effects on ad performance (e.g. permanent versus transitory).

In the empirical application, we find evidence of substantial persistent time variation in ad performance, emphasizing the importance of addressing dynamics in SEA ad performance models. We also find evidence of position and bidding endogeneity, indicating that purely predictive models are unable to capture causal relationships between ad position and clickthrough and conversion probabilities.

We note several limitations of this study. First, a drawback of the proposed method is the large computation time involved. Especially drawing the auxiliary Polya-Gamma variables is time-consuming, due to the large number of impressions and clicks in the dataset. Second, in the empirical application we use the cost-per-click (CPC) to proxy the bid. The actual bid may contain more information on the resulting ad position than the CPC. Finally, data could be missing not at random. For example, an advertiser might not bid on keywords that are expected to perform poorly. In this case, the results of the model might not hold for the non-selected keywords.

We note two interesting ways in which this study can be extended. First, one can add correlation across the time-varying parameters of the clickthrough and conversion equations. Such correlations can capture the idea that some shocks affect both clickthrough and conversion probabilities. On the downside, allowing for these correlations will substantially increase computation time as the time-varying parameter series for the two equations then need to be drawn jointly. Finally, one can use latent factors to explain ad performance instead of pre-specified keyword characteristics. Such an analysis will aid understanding of which factors drive the difference in ad performance across keywords and will help advertisers in designing effective ad campaigns. Again, this will substantially increase computation time.

References

Abou Nabout, N., Skiera, B., Stepanchuk, T., & Gerstmeier, E. (2012). An analysis of the profitability of fee-based compensation plans for search engine marketing. *International Journal of Research in Marketing*, *29*(1), 68–80.

Agarwal, A., Hosanagar, K., & Smith, M. D. (2011). Location, location, location: An analysis of profitability of position in online advertising markets. *Journal of Marketing Research*, *48*(6), 1057–1073.

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*(422), 669–679.

Blake, T., Nosko, C., & Tadelis, S. (2015). Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica*, *83*(1), 155–174.

Borgs, C., Chayes, J., Immorlica, N., Jain, K., Etesami, O., & Mahdian, M. (2007). Dynamics of bid optimization in online advertisement auctions. In *Proceedings of the 16th International Conference on World Wide Web* (pp. 531–540).

Carter, C. K., & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, *81*(3), 541–553.

Cary, M., Das, A., Edelman, B., Giotis, I., Heimerl, K., Karlin, A. R., ... Schwarz, M. (2007). Greedy bidding strategies for keyword auctions. In *Proceedings of the 8th*

ACM Conference on Electronic Commerce (pp. 262–271).

Dickey, J. M. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics*, 204–223.

Dinner, I. M., Van Heerde, H. J., & Neslin, S. A. (2014). Driving online and offline sales: The cross-channel effects of traditional, online display, and paid search advertising. *Journal of Marketing Research*, *51*(5), 527–545.

Durbin, J., & Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, *89*(3), 603–616.

Durbin, J., & Koopman, S. J. (2012). *Time Series Analysis by State Space Methods* (No. 38). Oxford University Press.

Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*(8), 1–18. Retrieved from `http://www.jstatsoft.org/v40/i08/` doi: 10.18637/jss.v040.i08

Edelman, B., Ostrovsky, M., & Schwarz, M. (2007). Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *The American Economic Review*, *97*(1), 242-259.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, *15*(2), 183–202.

Frühwirth-Schnatter, S., & Frühwirth, R. (2010). Data augmentation and MCMC for binary and multinomial logit models. In *Statistical Modelling and Regression Structures* (pp. 111–132). Springer. Available at ResearchGate DOI 10.1007/978-3-7908-2413-1_7.

Gelfand, A. E., & Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*(410), 398–409.

Geman, S., & Geman, D. (1987). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. In *Readings in Computer Vision* (pp. 564–584). Elsevier.

Ghose, A., & Yang, S. (2008). Analyzing search engine advertising: Firm behavior and cross-selling in electronic markets. In *Proceedings of the 17th International Conference on World Wide Web* (pp. 219–226).

Ghose, A., & Yang, S. (2009). An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science*, *55*(10), 1605–1622.

Graepel, T., Candela, J. Q., Borchert, T., & Herbrich, R. (2010). Web-scale Bayesian click-through rate prediction for sponsored search advertising in Microsoft's Bing search engine. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (pp. 13–20).

Greene, W. H. (2012). *Econometric Analysis* (seventh ed.). Pearson Education International.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton university press Princeton.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109.

Holmes, C. C., & Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, *1*(1), 145–168.

Jungbacker, B., & Koopman, S. J. (2015). Likelihood-based dynamic factor analysis for measurement and forecasting. *The Econometrics Journal*, *18*(2), C1–C21.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Li, H., & Kannan, P. (2014). Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research*, *51*(1), 40–56.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092.

Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association*, *108*(504), 1339–1349.

R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Rossi, P. E., Allenby, G. M., & McCulloch, R. (2005). *Bayesian Statistics and Marketing*. John Wiley & Sons.

Rutz, O. J., & Bucklin, R. E. (2007). A model of individual keyword performance in paid search advertising. *Available at SSRN 1024765*.

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

Rutz, O. J., Bucklin, R. E., & Sonnier, G. P. (2012). A latent instrumental variables approach to modeling keyword conversion in paid search advertising. *Journal of Marketing Research*, *49*(3), 306–319.

Rutz, O. J., Trusov, M., & Bucklin, R. E. (2011). Modeling indirect effects of paid search advertising: Which keywords lead to more future visits? *Marketing Science*, *30*(4), 646–665.

Ryan, D. (2016). *Understanding digital marketing: Marketing strategies for engaging the digital generation*. Kogan Page Publishers.

Skiera, B., & Abou Nabout, N. (2013). PROSAD: A bidding decision support system for profit optimizing search engine advertising. *Marketing Science*, *32*(2), 213–220.

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*(398), 528–540.

Van Heerde, H. J., Mela, C. F., & Manchanda, P. (2004). The dynamic effect of innovation on market structure. *Journal of Marketing Research*, *41*(2), 166–183.

Windle, J., Polson, N., & Scott, J. (2014a). BayesLogit: Bayesian logistic regression. *R package version: 0.5*, *1*.

Windle, J., Polson, N. G., & Scott, J. G. (2014b). Sampling Pólya-Gamma random variates: Alternate and approximate techniques. *arXiv preprint arXiv:1405.0506*.

Yang, S., & Ghose, A. (2010). Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marketing Science*, *29*(4), 602–623.

Yao, S., & Mela, C. F. (2011). A dynamic model of sponsored search advertising. *Marketing Science*, *30*(3), 447–468.

A. Gibbs sampler

To obtain posterior results of the dynamic Bayesian model, we use a Gibbs sampler with Polya-Gamma data augmentation (Geman & Geman, 1987, Tanner & Wong, 1987, Polson et al., 2013). The Polya-Gamma data augmentation scheme is suitable for binomial likelihoods (Polson et al., 2013). This scheme involves introducing one layer of auxiliary latent variables that follows a Polya-Gamma distribution. Conditional on these latent variables, the posterior distribution of the parameters of interest has the same functional form as the posterior distribution of parameters from a linear regression model with normally distributed error terms. This approach is similar to the data augmentation scheme for probit models of Albert and Chib (1993), but requires less memory storage as latent variables are drawn for each observation (keyword times day) instead of for each impression or click. In a SEA application this is crucial as the number of daily impressions and clicks can be very large.

The Polya-Gamma data augmentation scheme works as follows. Suppose that we have a binomially distributed variable $y \sim BIN(N, 1/(1+\exp(-\theta)))$. Introduce an auxiliary random variable ω that follows the Polya-Gamma distribution $PG(N, 0)$. The likelihood function $p(y|\theta)$ can be written as

$$p(y|\theta) = \left(\frac{1}{1+\exp(-\theta)}\right)^{y}\left(\frac{1}{1+\exp(\theta)}\right)^{N-y} = \int p(y|\theta,\omega)p(\omega)d\omega,$$

where Polson et al. (2013) have showed that the conditional distribution $p(y|\theta,\omega)$ is proportional to the likelihood kernel of a linear regression model

$$p(y|\theta,\omega) \propto \exp\left\{-\frac{\omega}{2}(z-\theta)^2\right\},$$

with pseudo-observations $z \equiv (y - N/2)/\omega$, signal θ, and independently distributed error terms with variances $1/\omega$. Thus, the full conditional distribution of θ, $p(\theta|\omega, y) \propto p(y|\theta,\omega)p(\theta)$, becomes standard. That is, the full conditional distribution of θ is the same as if we have the linear regression model $z = \theta + \nu$, $\nu \sim N(0, 1/\omega)$ with prior $p(\theta)$. Moreover, the full conditional distribution $p(\omega|\theta, y)$ is also a Polya-Gamma distribution, and ω can thus be easily sampled along in the Gibbs sampler (Polson et al., 2013).

For our dynamic model in Equations (1)-(4), we have that conditional on the auxiliary latent Polya-Gamma distributed variables for the clickthrough and conversion equations (denoted by ω_{it}^{CTR} and ω_{it}^{CON}, respectively) we have the multivariate linear regression model

$$z_{it}^{CTR} = \alpha_i^{CTR} + x_i'\beta_t^{CTR} + \lambda_i^{CTR}\ln(\mathrm{POS}_{it}) + s_t'\gamma^{CTR} + \eta_{it}^{CTR} + \xi_{it}^{CTR}, \tag{12}$$
$$z_{it}^{CON} = \alpha_i^{CON} + x_i'\beta_t^{CON} + \lambda_i^{CON}\ln(\mathrm{POS}_{it}) + s_t'\gamma^{CON} + \eta_{it}^{CON} + \xi_{it}^{CON}, \tag{13}$$
$$\ln(\mathrm{POS}_{it}) = \alpha_i^{POS} + x_i'\beta_t^{POS} + \lambda_i^{POS}\ln(\mathrm{BID}_{it}) + \psi^{POS}\ln(\mathrm{QS}_{it}) + s_t'\gamma^{POS} + \eta_{it}^{POS}, \tag{14}$$
$$\ln(\mathrm{BID}_{it}) = \alpha_i^{BID} + x_i'\beta_t^{BID} + q_{it}'\delta_i^{BID} + s_t'\gamma^{BID} + \eta_{it}^{BID}, \tag{15}$$

with $z_{it}^{CTR} \equiv (N_{it} - I_{it}/2)/\omega_{it}^{CTR}$, $z_{it}^{CON} \equiv (M_{it} - N_{it}/2)/\omega_{it}^{CON}$, $\xi_{it}^{CTR} \sim N(0, 1/\omega_{it}^{CTR})$, and $\xi_{it}^{CON} \sim N(0, 1/\omega_{it}^{CON})$. The equations are related through $\eta_{it} \sim MVN(0, \Sigma_\eta)$.

Now, for ease of representation, we replace the names CTR, CON, POS and BID by the numbers 1 to 4, respectively, and rename the variables and parameters to obtain the specific blocks for the Gibbs sampler. That is, we rewrite Equations (12)-(15) as

$$z_{1it} = w_{1it}'\pi_{1i} + x_i'\beta_{1t} + s_{1it}'\gamma_1 + \eta_{1it} + \xi_{1it}, \tag{16}$$
$$z_{2it} = w_{2it}'\pi_{2i} + x_i'\beta_{2t} + s_{2it}'\gamma_2 + \eta_{2it} + \xi_{2it}, \tag{17}$$
$$z_{3it} = w_{3it}'\pi_{3i} + x_i'\beta_{3t} + s_{3it}'\gamma_3 + \eta_{3it}, \tag{18}$$
$$z_{4it} = w_{4it}'\pi_{4i} + x_i'\beta_{4t} + s_{4it}'\gamma_4 + \eta_{4it}, \tag{19}$$

where

$$\begin{bmatrix} z_{1it} \\ z_{2it} \\ z_{3it} \\ z_{4it} \end{bmatrix} = \begin{bmatrix} z_{it}^{CTR} \\ z_{it}^{CON} \\ \ln(\mathrm{POS}_{it}) \\ \ln(\mathrm{BID}_{it}) \end{bmatrix}, \quad \begin{bmatrix} w'_{1it} \\ w'_{2it} \\ w'_{3it} \\ w'_{4it} \end{bmatrix} = \begin{bmatrix} 1 & \ln(\mathrm{POS}_{it}) \\ 1 & \ln(\mathrm{POS}_{it}) \\ 1 & \ln(\mathrm{BID}_{it}) \\ 1 & q'_{it} \end{bmatrix}, \quad \begin{bmatrix} \pi'_{1i} \\ \pi'_{2i} \\ \pi'_{3i} \\ \pi'_{4i} \end{bmatrix} = \begin{bmatrix} \alpha_i^{CTR} & \lambda_i^{CTR} \\ \alpha_i^{CON} & \lambda_i^{CON} \\ \alpha_i^{POS} & \lambda_i^{POS} \\ \alpha_i^{BID} & \delta_i'^{BID} \end{bmatrix}, \quad \begin{bmatrix} \beta_{1t} \\ \beta_{2t} \\ \beta_{3t} \\ \beta_{4t} \end{bmatrix} = \begin{bmatrix} \beta_t^{CTR} \\ \beta_t^{CON} \\ \beta_t^{POS} \\ \beta_t^{BID} \end{bmatrix},$$

and

$$\begin{bmatrix} s'_{1it} \\ s'_{2it} \\ s'_{3it} \\ s'_{4it} \end{bmatrix} = \begin{bmatrix} s'_t & \\ s'_t & \\ s'_t & \ln(QS_{it}) \\ s'_t & \end{bmatrix}, \quad \begin{bmatrix} \gamma'_1 \\ \gamma'_2 \\ \gamma'_3 \\ \gamma'_4 \end{bmatrix} = \begin{bmatrix} \gamma'^{CTR} & \\ \gamma'^{CON} & \\ \gamma'^{POS} & \psi^{POS} \\ \gamma'^{BID} & \end{bmatrix}, \quad \begin{bmatrix} \eta_{1it} \\ \eta_{2it} \\ \eta_{3it} \\ \eta_{4it} \end{bmatrix} = \begin{bmatrix} \eta_{it}^{CTR} \\ \eta_{it}^{CON} \\ \eta_{it}^{POS} \\ \eta_{it}^{BID} \end{bmatrix}, \quad \begin{bmatrix} \xi_{1it} \\ \xi_{2it} \end{bmatrix} = \begin{bmatrix} \xi_{it}^{CTR} \\ \xi_{it}^{CON} \end{bmatrix}.$$

Equations (6)-(9) are also rewritten in terms of j:

$$\begin{aligned} &\beta_{j,t+1} = \Phi_j \beta_{jt} + \nu_{jt}, \quad \nu_{jt} \sim MVN(0, \Sigma_{\beta,j}), \quad \beta_{j1} \sim MVN(0, 5\Sigma_{\beta,j}), && \text{for } j = 1, ..., 4, \\ &\alpha_{ji} \sim N(x'_i \tilde{\alpha}_j, \sigma^2_{\alpha,j}), && \text{for } j = 1, ..., 4, \\ &\lambda_{ji} \sim N(x'_i \tilde{\lambda}_j, \sigma^2_{\lambda,j}), && \text{for } j = 1, ..., 3, \\ &\delta_{ji} \sim MVN(x'_i \tilde{\delta}_j, \Sigma_{\delta,j}), && \text{for } j = 4. \end{aligned}$$

For computational efficiency, in the Gibbs sampler we draw the parameters of each of the four model equations *separately* by conditioning on the η_{it} of the other equations. This also helps deal with missing values. For this, we compute

$$\begin{aligned} \bar{\eta}_{jit} &\equiv \mathrm{E}[\eta_{jit} | \eta_{-j,it}] &&= \Sigma_{\eta(j,-j)} \Sigma^{-1}_{\eta(-j,-j)} \eta_{-j,it}, \\ \bar{\sigma}^2_{\eta,j} &\equiv \mathrm{Var}(\eta_{jit} | \eta_{-j,it}) &&= \Sigma_{\eta(j,j)} - \Sigma_{\eta(j,-j)} \Sigma^{-1}_{\eta(-j,-j)} \Sigma_{\eta(-j,j)}, \end{aligned}$$

for $j = 1, ..., 4$. Here we denote by $\eta_{-j,it}$ all elements in η_{it} except for the j^{th} element and any missing elements, and by $\Sigma_{\eta(j,-j)}$ all elements of Σ_η related to row j and all columns except for the j^{th}. Then, we can rewrite each of the Equations (16)-(19) as a univariate regression model conditional on $\eta_{-j,it}$, as given by

$$z_{jit} = w'_{jit} \pi_{ji} + x'_i \beta_{jt} + s'_{jit} \gamma_j + \bar{\eta}_{jit} + \zeta_{jit}, \tag{20}$$

for $j = 1, ..., 4$, where the introduction of $\bar{\eta}_{jit}$ ensures that the error terms ζ_{jit} are independent of each other

$$\begin{pmatrix} \zeta_{1it} \\ \zeta_{2it} \\ \zeta_{3it} \\ \zeta_{4it} \end{pmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\zeta,1it} \equiv 1/\omega_{1it} + \bar{\sigma}^2_{\eta,1} & 0 & 0 & 0 \\ 0 & \sigma^2_{\zeta,2it} \equiv 1/\omega_{2it} + \bar{\sigma}^2_{\eta,2} & 0 & 0 \\ 0 & 0 & \sigma^2_{\zeta,3it} \equiv \bar{\sigma}^2_{\eta,3} & 0 \\ 0 & 0 & 0 & \sigma^2_{\zeta,4it} \equiv \bar{\sigma}^2_{\eta,4} \end{bmatrix} \right).$$

Hence, after drawing the parameters of a single equation j, we update the η_j of that equation, and condition on the new η_j in drawing the parameters of the next equations.

A.1. Overview Gibbs sampler

We use the rewritten model in Equation (20) to construct the Gibbs sampler. To improve the mixing rates of the sampler, we (i) jointly sample η_{jit} and the parameters in (α_{ji}, β_{jt}, λ_{ji}, δ_{ji}, γ_j), and (ii) jointly sample η and Σ_η. Because of sampling jointly, we need to draw η twice within a single Gibbs step.

The Gibbs steps are given by

1. For $j = 1, ..., 4$ do

   i. Sample $\omega_{jit}|\ I_{it},\ N_{it},\ M_{it},\ \pi_{ji},\ \beta_{jt},\ \gamma_j,\ \eta_{jit}$ (if $j \in \{1, 2\}$, for $i = 1, ..., N$, $t = 1, ..., T$).

   ii. Compute $z_{jit}|\ \omega_{jit},\ I_{it},\ N_{it},\ M_{it}$ (if $j \in \{1, 2\}$, for $i = 1, ..., N$, $t = 1, ..., T$).

   iii. Compute $\bar{\eta}_{jit},\ \bar{\sigma}^2_{\eta,j}|\ \eta_{-j,it},\ \Sigma_\eta$ (for $i = 1, ..., N$ and $t = 1, ..., T$).

   iv. Sample $\pi_{ji} = (\alpha_{ji}, \lambda_{ji}, \delta_{ji})|\ z_{ji},\ \omega_{ji},\ \beta_j,\ \gamma_j,\ \tilde{\alpha}_j,\ \tilde{\lambda}_j,\ \tilde{\delta}_j,\ \bar{\eta}_{ji},\ \sigma^2_{\alpha,j},\ \sigma^2_{\lambda,j},\ \Sigma_{\delta,j},\ \bar{\sigma}^2_{\eta,j}$ (for $i = 1, ..., N$).

   v. Sample $\{\beta_{jt}\}_{t=1}^T|\ z_j,\ \omega_j,\ \pi_j,\ \gamma_j,\ \bar{\eta}_j,\ \Sigma_{\beta,j},\ \Phi_j,\ \bar{\sigma}^2_{\eta,j}$.

   vi. Sample $\gamma_j|\ z_j,\ \omega_j,\ \pi_j,\ \beta_j,\ \bar{\eta}_j,\ \bar{\sigma}^2_{\eta,j}$.

   vii. Sample η_{jit} (for $i = 1, ..., N$ and $t = 1, ..., T$):

      a. If $j \in \{1, 2\}$, sample $\eta_{jit}|\ z_{jit},\ \omega_{jit},\ \pi_{ji},\ \beta_{jt},\ \gamma_j,\ \eta_{-j,it},\ \Sigma_\eta$.

      b. If $j \in \{3, 4\}$, compute $\eta_{jit}|\ z_{jit},\ \pi_{ji},\ \beta_{jt},\ \gamma_j$.

2. Sample $\tilde{\alpha}|\ \alpha,\ \Sigma_\alpha$, sample $\tilde{\lambda}|\ \lambda,\ \Sigma_\lambda$, and sample $\tilde{\delta}|\ \delta,\ \Sigma_\delta$.

3. Sample $\Sigma_\alpha|\ \alpha,\ \tilde{\alpha}$, sample $\Sigma_\lambda|\ \lambda,\ \tilde{\lambda}$, and sample $\Sigma_\delta|\ \delta,\ \tilde{\delta}$.

4. Sample $\Phi|\ \beta,\ \Sigma_\beta$.

5. Sample $\Sigma_\beta|\ \beta,\ \Phi$.

6. Sample $\Sigma_\eta | \; z, \omega, \pi, \beta, \gamma$.
7. Compute $\bar{\eta}_{jit}, \bar{\sigma}^2_{\eta,j} | \; \eta_{-j,it}, \Sigma_\eta$ (for $j = 1, ..., 4$, $i = 1, ..., N$ and $t = 1, ..., T$).
8. Sample $\eta_{jit} | \; z_{jit}, \omega_{jit}, \pi_{ji}, \beta_{jt}, \gamma_j, \eta_{-j,it}, \Sigma_\eta$ (for $j = 1, 2$, $i = 1, ..., N$ and $t = 1, ..., T$).

A.2. Priors

We choose conjugate priors to ensure that the model parameters can be drawn using Gibbs steps. For the logit equations (clickthrough and conversion) we take slightly informative priors, for the normal equations (position and bid) we take rather uninformative priors. First, for the means of the keyword-specific parameters ($\tilde{\alpha}$, $\tilde{\lambda}$, and $\tilde{\delta}$) we take multivariate normal prior distributions with mean 0 and covariance matrix I (for clickthrough and conversion equations) or $100I$ (for position and bid equations). Second, for the variances of the keyword-specific parameters (diagonal elements of Σ_α, Σ_λ, and Σ_δ) we take inverse Gamma-2 prior distributions with shape parameter $\kappa_0 = 5$ and scale parameter $\kappa_1 = 5 \times 0.1$.

Third, for the time-varying parameters we take a multivariate normal prior for Φ with mean $\hat{\Phi}_0 = 0.5\iota$ and covariance matrix $\Sigma_{\Phi_0} = 0.5I$, where ι represents a vector of ones. For the diagonal elements of Σ_β we take inverse Gamma-2 prior distributions with shape parameter $\kappa_{\beta,0} = 5$ and scale parameter $\kappa_{\beta,1} = 5 \times 0.001$. Fourth, for the time-invariant parameters in (γ, ψ^{POS}) we take a multivariate normal prior distribution with mean 0 and covariance matrix I (for clickthrough and conversion equations) or $100I$ (for position and bid equations).

Finally, for the covariance matrix Σ_η we take an inverse Wishart prior with 8 degrees of freedom and scale matrix $8 \times 0.1I$.

A.3. Initialization

We take the following initialization. For the baseline level, we set $\Sigma_\alpha = 0.1I$ and $\tilde{\alpha} = 0$ except for the intercept in $\tilde{\alpha}$ which we set to -3 for the clickthrough and conversion equations, to 1 for the position equation, and to -1 for the bid equation. For the time-varying parameters, we set $\{\beta_t\}_{t=1}^T = 0$, $\Phi = 0.5I$, and $\Sigma_\beta = 0.001I$. Furthermore, we initialize $\gamma = 0$, $\psi^{POS} = 0$, $\tilde{\lambda} = 0$, and $\Sigma_\lambda = 0.1I$. For the instruments, we set $\tilde{\delta} = 0$ and $\Sigma_\delta = 0.1I$. Finally, for the keyword- and time-specific shocks, we initialize $\Sigma_\eta = 0.1I$, η_{it}^{CTR} and η_{it}^{CON} to 0 for all i and t and compute η_{it}^{POS} and η_{it}^{CPC} based on the other initializations.

A.4. Steps Gibbs sampler

A.4.1. Sampling Polya-Gamma variables ω

The full conditional posterior distribution of the auxiliary latent Polya-Gamma variables ω_{1it} (ω_{2it}) are independent Polya-Gamma distributions with parameters I_{it} (N_{it}) and θ_{1it} (θ_{2it}) for $i = 1, ..., N$ and $t = 1, ..., T$ where I_{it} (N_{it}) denotes the number impressions (clicks), and

$$\theta_{jit} = w'_{jit}\pi_{ji} + x'_i\beta_{jt} + s'_t\gamma_j + \eta_{jit},$$

for $j = 1, 2$. We draw the Polya-Gamma variables using the R package BayesLogit (Windle, Polson, & Scott, 2014a). For computational efficiency, we approximate the Polya-Gamma variable ω_{1it} (ω_{2it}) by normal variables in case $I_{it} > 170$ ($N_{it} > 170$) (Windle, Polson, & Scott, 2014b). In this approximation, we set the first and second moment of the normal distribution equal to the first and second moment of the associated Polya-Gamma distribution.

After drawing the Polya-Gamma variables, we compute the pseudo data points for the clickthrough and conversion equations

$$\begin{aligned} z_{1it} &= \left(N_{it} - I_{it}/2\right)/\omega_{1it}, \\ z_{2it} &= \left(M_{it} - N_{it}/2\right)/\omega_{2it}. \end{aligned}$$

A.4.2. Sampling α_i, λ_i, and δ_i

To sample α_{ji}, λ_{ji}, and δ_{ji} (collected in π_{ji}) for $i = 1, ..., N$, note that we can write Equation (20) as the univariate normal regression model

$$y_{\pi,jit} \equiv z_{jit} - x'_i\beta_{jt} - s'_{jit}\gamma_j - \bar{\eta}_{jit} = w'_{jit}\pi_{ji} + \zeta_{jit}, \quad \zeta_{jit} \sim N(0, \sigma^2_{\zeta,jit}),$$

for $t = 1, ..., T$, with a normal prior for $\pi_{ji} \sim MVN(\bar{\pi}_{j0}, \Sigma_{\pi_{j0}})$ where

$$\bar{\pi}_{j0} = \begin{bmatrix} x'_i\tilde{\alpha}_j \\ x'_i\tilde{\lambda}_j \end{bmatrix}, \quad \Sigma_{\pi_{j0}} = \begin{bmatrix} \sigma^2_{\alpha,j} & 0 \\ 0 & \sigma^2_{\lambda,j}, \end{bmatrix}.$$

When $j = 4$, the elements $\tilde{\lambda}_j$ and $\sigma^2_{\lambda,j}$ are replaced by $\tilde{\delta}_j$ and $\Sigma_{\delta,j}$.

We draw π_{ji} from $MVN(\hat{\pi}_{ji}, \hat{\Sigma}_{\pi_{ji}})$

$$\hat{\Sigma}_{\pi_{ji}} = \left(\sum\nolimits_{t=1}^{T} w_{jit}w'_{jit}/\sigma^2_{\zeta,jit} + \Sigma^{-1}_{\pi_{j0}}\right)^{-1}, \qquad \hat{\pi}_{ji} = \hat{\Sigma}_{\pi_{ji}}\left(\sum\nolimits_{t=1}^{T} w_{jit}y_{\pi,jit}/\sigma^2_{\zeta,jit} + \Sigma^{-1}_{\pi_{j0}}\bar{\pi}_0\right),$$

for $i = 1, ..., N$.

A.4.3. Sampling β_t

To sample $\{\beta_{jt}\}_{t=1}^{T}$, note that we can write Equation (20) as the univariate normal regression model

$$y_{\beta,jit} \equiv z_{jit} - w'_{jit}\pi_{ji} - s'_{jit}\gamma_j - \bar{\eta}_{jit} = x'_i\beta_{jt} + \zeta_{jit}, \quad \zeta_{jit} \sim N(0, \sigma^2_{\zeta,jit}),$$

with

$$\beta_{j,t+1} = \Phi_j\beta_{jt} + \nu_{jt}, \qquad \nu_{jt} \sim MVN(0, \Sigma_{\beta,j}), \qquad \beta_{j1} \sim MVN(0, 5\Sigma_{\beta,j}).$$

We sample $\{\beta_{jt}\}_{t=1}^{T}$ using the simulation smoother of Durbin and Koopman (2002) (as explained in Durbin and Koopman (2012), section 4.9.2). To speed up computations, we perform collapsed filtering (Durbin and Koopman 2012, Chapter 6.5, Jungbacker and Koopman 2015).

Collapsed filtering works as follows. We have the $(N \times 1)$ vector $y_{\beta,jt}$ and the $(N \times K)$ matrix X, with $K >> N$. We can compute a $(K \times N)$ matrix A^*_{jt} such that we can obtain the correct smoothed estimates by using the observation equation with $(K \times 1)$ observation vector

$$A^*_{jt}y_{\beta,jt} = A^*_{jt}X + A^*_{jt}\zeta_{jt},$$

where the covariance matrix of ζ_{jt} is a diagonal matrix with elements $\sigma^2_{\zeta,jit}$. Hence, this procedure allows for much lower computation times because the altered observation vector is of much smaller dimension than the original observation vector, while the covariance matrix remains diagonal. We take the i^{th} column of A^*_{jt} equal to

$$A^*_{jit} = \begin{cases} \left(\sum_{n:I_{nt}\geq 1}^{N} \frac{1}{\sigma^2_{\zeta,jnt}} x_n x'_n\right)^{-1/2} x_i/\sigma^2_{\zeta,jit}, & \text{if } j = 1, 3, \\ \left(\sum_{n:N_{nt}\geq 1}^{N} \frac{1}{\sigma^2_{\zeta,jnt}} x_n x'_n\right)^{-1/2} x_i/\sigma^2_{\zeta,jit}, & \text{if } j = 2, 4, \end{cases}$$

where for the matrix in the first terms on the right hand sides we first take the Cholesky decomposition (upper triangular) and then the inverse. This A^* is chosen because then $A^*_{jt}\zeta_{jt} \sim MVN(0, I)$.

In case X is not of full column rank (see Section 4.2) the above procedure needs to be slightly altered. That is, let $\tilde{X}$ be the $(N \times K_2)$ matrix with columns of X such that $\tilde{X}$ is of full column rank and has the same columnn space as X. Then, we compute A^*_{jt} using the rows of $\tilde{X}$ instead of X.

Finally, we have to deal with missing values, which in the collapsed case refers to time periods in which there is a variable k^* in x_i which has the same value over all i. In other words, that time period has no keywords with impressions/clicks that have a specific characteristic in x_i. When such missings occur, we set element k^* in $A^*_{jt} y_{\beta,jt}$ to zero, and the elements in the $k^{*(th)}$ row and column of $A^*_{jt} X$ equal to zero.

A.4.4. Sampling γ

To sample γ_j note that we can write Equation (20) as the univariate normal regression model

$$y_{\gamma,jit} \equiv z_{jit} - w'_{jit}\pi_{ji} - x'_i\beta_{jt} - \bar{\eta}_{jit} = s'_{jit}\gamma_j + \zeta_{jit}, \quad \zeta_{jit} \sim N(0, \sigma^2_{\zeta,jit}).$$

We draw γ_j from $MVN(\hat{\gamma}_j, \hat{\Sigma}_{\gamma_j})$ where

$$\hat{\Sigma}_{\gamma_j} = \left(\sum\nolimits_{i=1}^{N} \sum\nolimits_{t=1}^{T} s_{jit}s'_{jit}/\sigma^2_{\zeta,it} + \Sigma^{-1}_{\gamma_{j0}}\right)^{-1}, \qquad \hat{\gamma}_j = \hat{\Sigma}_{\gamma_j}\left(\sum\nolimits_{i=1}^{N} \sum\nolimits_{t=1}^{T} s_{jit}y_{\gamma,jit}/\sigma^2_{\zeta,it}\right),$$

where $\Sigma_{\gamma_{j0}}$ is the diagonal covariance matrix of the normal prior for γ.

A.4.5. Sampling η

Next we sample $\{\{\eta_{jit}\}_{i=1}^N\}_{t=1}^T$. In case $j \in \{3, 4\}$ (position and bid equations), we see from Equations (18) and (19) that we can directly compute η_{jit}:

$$\eta_{jit} = z_{jit} - w'_{jit}\pi_{ji} - x'_i\beta_{jt} - s'_{jit}\gamma_j. \tag{21}$$

In case $j \in \{1, 2\}$ (clickthrough and conversion equations), we sample η_{jit}. Note that we can write both Equations (16) and (17) as the univariate normal regression model

$$y_{\eta,jit} \equiv z_{jit} - w'_{jit}\pi_{ji} - x'_i\beta_{jt} - s'_{jit}\gamma_j = \eta_{jit} + \xi_{jit}, \quad \xi_{jit} \sim N(0, 1/\omega_{jit}),$$

with a normal prior $\eta_{jit} \sim N(\bar{\eta}_{jit}, \bar{\sigma}^2_{\eta,j})$. We draw η_{jit} for $i = 1, ..., N$ and $t = 1, ..., T$ from $N(\hat{\eta}_{jit}, \hat{\Sigma}_{\eta_{jit}})$ where

$$\hat{\Sigma}_{\eta_{jit}} = \left(\omega_{jit} + 1/\bar{\sigma}^2_{\eta,j}\right)^{-1}, \qquad \hat{\eta}_{jit} = \hat{\Sigma}_{\eta_{jit}}\left(\omega_{jit}y_{\eta,jit} + \bar{\eta}_{jit}/\bar{\sigma}^2_{\eta,j}\right).$$

A.4.6. Sampling $\tilde{\alpha}$, $\tilde{\lambda}$, and $\tilde{\delta}$

To sample $\tilde{\alpha}$, note that Equation (7) is a multivariate regression model given $\{\alpha_i\}_{i=1}^N$ and Σ_α. We draw $\tilde{\alpha}$ from $MVN(\hat{\tilde{\alpha}}, \hat{\Sigma}_{\tilde{\alpha}})$ where

$$\hat{\Sigma}_{\tilde{\alpha}} = \left(\sum\nolimits_{i=1}^N (I_4 \otimes x_i')'\Sigma_\alpha^{-1}(I_4 \otimes x_i') + \Sigma_{\tilde{\alpha}_0}^{-1}\right)^{-1}, \qquad \hat{\tilde{\alpha}} = \hat{\Sigma}_{\tilde{\alpha}} \left(\sum\nolimits_{i=1}^N (I_4 \otimes x_i')'\Sigma_\alpha^{-1}\alpha_i\right),$$

where $\Sigma_{\tilde{\alpha}_0}$ is the covariance matrix of the normal prior for $\tilde{\alpha}$, and $\otimes$ denotes the Kronecker product.

To sample $\tilde{\lambda}$, note that Equation (8) is a multivariate regression model given $\{\lambda_i\}_{i=1}^N$ and Σ_λ. We draw $\tilde{\lambda}$ from $MVN(\hat{\tilde{\lambda}}, \hat{\Sigma}_{\tilde{\lambda}})$ where

$$\hat{\Sigma}_{\tilde{\lambda}} = \left(\sum\nolimits_{i=1}^N (I_3 \otimes x_i')'\Sigma_\lambda^{-1}(I_3 \otimes x_i') + \Sigma_{\tilde{\lambda}_0}^{-1}\right)^{-1}, \qquad \hat{\tilde{\lambda}} = \hat{\Sigma}_{\tilde{\lambda}} \left(\sum\nolimits_{i=1}^N (I_3 \otimes x_i')'\Sigma_\lambda^{-1}\lambda_i\right),$$

where $\Sigma_{\tilde{\lambda}_0}$ is the covariance matrix of the normal prior for $\tilde{\lambda}$.

To sample $\tilde{\delta}$, note that Equation (9) is a multivariate regression model given $\{\delta_i\}_{i=1}^N$ and Σ_δ. We draw $\tilde{\delta}$ from $MVN(\hat{\tilde{\delta}}, \hat{\Sigma}_{\tilde{\delta}})$ where

$$\hat{\Sigma}_{\tilde{\delta}} = \left(\sum\nolimits_{i=1}^N \Sigma_\delta^{-1} + \Sigma_{\tilde{\delta}_0}^{-1}\right)^{-1}, \qquad \hat{\tilde{\delta}} = \hat{\Sigma}_{\tilde{\delta}} \left(\sum\nolimits_{i=1}^N \Sigma_\delta^{-1}\delta_i\right),$$

where $\Sigma_{\tilde{\delta}_0}$ is the covariance matrix of the normal prior for $\tilde{\delta}$.

A.4.7. Sampling Σ_α, Σ_λ, and Σ_δ

To sample Σ_α, Σ_λ, and $\Sigma_{\delta,j}$, note that these covariance matrices are diagonal. Therefore, we separately draw each diagonal element.

To sample $\sigma_{\alpha,j}^2$, note we have a univariate regression model for $\alpha_{ji} \sim N(x_i'\tilde{\alpha}_j, \sigma_{\alpha,j}^2)$ for $i = 1, ..., N$. We therefore draw $\sigma_{\alpha,j}^2$ from the inverse Gamma distribution

$$\sigma_{\alpha,j}^2 \sim IG\left(\frac{\sum_{i=1}^N (\alpha_{ji} - x_i'\tilde{\alpha}_j)^2 + \kappa_1}{2}, \frac{N + \kappa_0}{2}\right),$$

for $j = 1, ..., 4$, with prior parameters κ_0 and κ_1.

To sample $\sigma_{\lambda,j}^2$, note we have a univariate regression model for $\lambda_{ji} \sim N(x_i'\tilde{\lambda}_j, \sigma_{\lambda,j}^2)$ for

$i = 1, ..., N$. We therefore draw $\sigma^2_{\lambda,j}$ from the inverse Gamma distribution

$$\sigma^2_{\lambda,j} \sim IG\left(\frac{\sum_{i=1}^{N}(\lambda_{ji} - x_i'\tilde{\lambda}_j)^2 + \kappa_1}{2}, \frac{N + \kappa_0}{2}\right),$$

for $j = 1, ..., 3$, with prior parameters κ_0 and κ_1.

To sample $\Sigma_{\delta,j,kk}$, for $j = 4$, note we have a univariate regression model for $\delta_{ki} \sim N(x_i'\tilde{\delta}_{jk}, \Sigma_{\delta,j,kk})$ for $i = 1, ..., N$. We therefore draw $\Sigma_{\delta,kk}$ from the inverse Gamma distribution

$$\Sigma_{\delta,j,kk} \sim IG\left(\frac{\sum_{i=1}^{N}(\delta_{jki} - \tilde{\delta}_{jk})^2 + \kappa_1}{2}, \frac{N + \kappa_0}{2}\right),$$

for $k = 1, 2$, with prior parameters κ_0 and κ_1.

A.4.8. Sampling Φ

To sample Φ, note that Equation (6) is a multivariate regression model given β and Σ_β. We draw Φ from $MVN(\hat{\Phi}, \hat{\Sigma}_\Phi)$ where

$$\begin{aligned}\hat{\Sigma}_\Phi &= \left(\sum\nolimits_{t=2}^{T} \beta_{t-1}\Sigma_\beta^{-1}\beta_{t-1} + \Sigma_{\Phi_0}^{-1}\right)^{-1},\\ \hat{\Phi} &= \hat{\Sigma}_\Phi\left(\sum\nolimits_{t=2}^{T} \beta_{t-1}\Sigma_\beta^{-1}\beta_t + \Sigma_{\Phi_0}^{-1}\hat{\Phi}_0\right),\end{aligned}$$

where $\hat{\Phi}_0$ is the mean vector and Σ_{Φ_0} is the covariance matrix of the normal prior for Φ.

A.4.9. Sampling Σ_β

To sample Σ_β, note that Φ is a diagonal matrix and we can therefore draw each k^{th} diagonal elements separately. For the k^{th} element, we have the univariate regression model

$$\beta_{k,t+1} = \Phi_{kk}\beta_{kt} + \nu_{kt}, \qquad \nu_{kt} \sim N(0, \Sigma_{\beta,kk}), \qquad \beta_{k1} \sim N(0, 5\Sigma_{\beta,kk}).$$

We therefore draw $\Sigma_{\beta,kk}$ from the inverse Gamma distribution

$$\Sigma_{\beta,kk} \sim IG\left(\frac{\sum_{t=2}^{T}(\beta_{kt} - \Phi_{kk}\beta_{k,t-1})^2 + \beta_{k1}^2/5 + \kappa_{\beta,1}}{2}, \frac{T + \kappa_{\beta,0}}{2}\right),$$

with prior parameters $\kappa_{\beta,0}$ and $\kappa_{\beta,1}$.

A.4.10. Sampling Σ_η

To sample Σ_η in a computationally efficiently manner, we use an independence Metropolis-Hastings (MH) step (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953, Hastings, 1970). For this purpose, we first reparameterize Σ_η into elements that are unconstrained, using a Cholesky decomposition. Next, we draw a candidate for the unconstrained parameters from a multivariate normal distribution with as mean the posterior mode, and as covariance matrix the negative of the inverse of the Hessian of the log posterior at the posterior mode. To find the posterior mode and Hessian, we perform an optimization using the analytic gradient and an approximated Hessian from the outer-product-of-gradients (BHHH) method. Details are in the Supplementary Materials.[12]

[12] Available upon request.