Appendix A. The Markov Chain Monte Carlo Procedure

We follow Ghose and Yang (2009) in the estimation and employ Gibbs sampling in the Markov Chain Monte Carlo (MCMC) procedure, where variance-covariance matrix Σ and parameter vector β are drawn from their posterior distributions, conditional on each other. In the paper, we set the total number of iterations to 10,000 and used the first 5,000 iterations as burn-in.

Equations (1) to (4) as presented in the paper define a system of equations. We first define the variance-covariance matrix between all error terms in four equations as Σ with

$$\Sigma = \begin{bmatrix} v_\varepsilon^2 & & & \\ v_{\sigma\varepsilon} & v_\sigma^2 & & \\ v_{\omega\varepsilon} & v_{\omega\sigma} & v_\omega^2 & \\ v_{\nu\varepsilon} & v_{\nu\sigma} & v_{\nu\omega} & v_\nu^2 \end{bmatrix}.$$

We further define the matrix of coefficients for all equations as β with

$$\beta = \begin{bmatrix} \alpha & 0 & 0 & 0 \\ 0 & \varsigma & 0 & 0 \\ 0 & 0 & \gamma & 0 \\ 0 & 0 & 0 & \varphi \end{bmatrix}.$$

Since the model includes lagged and simultaneous variables from other equations in the system of four equations, we will have to treat these lagged and simultaneous variables as endogenous variables. The matrix of exogenous independent variables, is denoted as Z, which is common across four equations, including *Intercept*, *Length*, *Competing*, *Focal*, *Broad*, *Phrase*, and the interaction terms between *Competing* (*Focal*) and other exogenous independent variables (*Focal* and *Competing* excluded). The matrices of endogenous independent variables are denoted as $\dot{X}_1$ to $\dot{X}_4$ for four equations, respectively. The data matrix is then denoted by $X = [X_1\ X_2\ X_3\ X_4]$ with $X_1 = [Z\ \ \dot{X}_1]$, $X_2 = [Z\ \ \dot{X}_2]$, $X_3 = [Z\ \ \dot{X}_3]$, $X_4 = [Z\ \ \dot{X}_4]$. We further denote β_Z as the coefficient matrix for exogenous variables and $\beta_{\dot{X}}$ for endogenous variables. The matrix of dependent variable is denoted as

$$Y = [CTR \quad CR \quad \log(CPC) \quad \log(Rank)].$$

The likelihood function, given the derived click-through rate (*CTR*) and conversion rate (*CR*), is as follows:

$$\mathcal{L}(Y, X|\beta, \Sigma) = \prod_i \phi(L_\Sigma^{-1}(y_i - X_i\beta)).$$

Where ϕ is the p.d.f. of a Standard Normal distribution and L_Σ^{-1} is the Cholesky decomposition of the variance-covariance matrix Σ.

To address the endogeneity problem, we adopt the well-known "orthogonalization" technique (Kim 2008, Hamilton 1994). In this setting, the endogeneity in the system of equations solely comes from the lagged and simultaneous variables. So the correlation between endogenous variables and the error terms is fully attributed to the correlation between the error terms. In other words, if the error terms in the four equations are not correlated, we would not have the endogeneity problem in X_1 to X_4 (Kim 2008). Given this special feature of the model, we can use the Cholesky decomposition to eliminate the correlations in the error terms. With a consistent estimate of Σ, we can multiply each side of the equations by the inverse of the Cholesky decomposition of Σ and resolve the endogeneity issue.

In our application, we use the following MCMC procedure to obtain consistent estimates of the coefficients as well as the variance-covariance matrix.

The initial values of β^0 are drawn from uninformative priors of the normal distribution with mean 0 and the variance-covariance matrix $100I_k$, where I_k is an identity matrix. The initial values of Σ^0 are drawn from an Inverse-Wishart distribution with scale matrix Q_0, an identity matrix, and with a degree-of-freedom, v_0, equal to 4. Inverse-Wishart distribution is desirable in this setting because it is the conjugate prior for covariance matrices of normally distributed variables.

In iteration *t* of the MCMC procedure, draw β^{t+1} and Σ^{t+1} from the following posterior distributions:

Step 1: Draw $\Sigma^{t+1}|\beta^{t}$ from an Inverse-Wishart distribution with $\Sigma^{t+1} \sim IW(Q_0 + e_t' e_t, v_0 + 4N)$. In the equation, $e_t = Y - (I_4 \otimes X) \times \beta_Z$ with β_Z a $K_1 \times 4$ matrix constructed from β^t with only parameters of variables in Z put into each column (K_1 is the dimension of the exogenous independent variables Z). I_4 is a 4×4 identity matrix and $\otimes$ is the Kronecker product. Q_0 and v_0 are prior hyper-parameters set to the identity matrix and 4. The reason we only use β_Z is that this way we can obtain consistent estimates of the parameters so we can use Cholesky decomposition to resolve the endogeneity problem. This step will give us a consistent estimate of the variance-covariance matrix (Kim 2008, Hamilton 1994). Notably, we only include exogenous variables in this step as the inclusion of endogenous variables will lead to an inconsistent estimation of Σ.

Step 2: Draw $\beta^{t+1}|\Sigma^{t}$ from a normal distribution with $\beta^{t+1} \sim Norm(\mu^{t+1}, \Omega^{t+1})$. In the distribution of β^{t+1}, $\mu^{t+1} = \Omega^{t+1}\tilde{X}'\tilde{Y}$ and $\Omega^{t+1} = (\tilde{X}'\tilde{X} + .01 I_k)^{-1}$ are posterior mean and posterior variance-covariance matrix for β^{t+1}. $\tilde{Y}$ is a $4N \times 1$ vector created by stacking the dependent variables of all keywords $\tilde{Y}_i = L_{\Sigma}^{-1} Y_i$ and $LL' = \Sigma^{t+1}$. $\tilde{X}$ is a $4N \times K$ matrix created by stacking the independent variables of all keywords, each taking the form

$$\tilde{X}_i = L^{-1} \times \begin{bmatrix} X_{1i} & 0 & 0 & 0 \\ 0 & X_{2i} & 0 & 0 \\ 0 & 0 & X_{3i} & 0 \\ 0 & 0 & 0 & X_{4i} \end{bmatrix}_{4 \times K}.$$

Given the independence assumption of the variables, the posterior distributions of the variance-covariance matrix Σ are Inverse-Wishart too.

With our Bayesian framework, we can iteratively update the parameters β and the variance-covariance matrix Σ. This has several advantages over traditional estimation approaches, such as maximum likelihood estimations (MLEs). First, the method is able to deal with large and sparse matrices. When traditional estimation methods cannot converge, Bayesian methods have the desirable feature of convergence with even sparse data (Rossi and Allenby 2003). Second, our method does random draws on the whole parameter space. After each draw, we are able to calculate the probability of observing the randomly drawn parameter data, conditional on the data available. This procedure allows us to explore information available on the whole surface of the likelihood/posterior, thus avoiding being trapped in local maxima.

References

C.-J. Kim. 2008. Dealing with Endogeneity in Regression Models with Dynamic Coefficients. *Foundations and Trends in Econometrics*, **3**(3), 165-266.

Hamilton, J. D. 1994. Time Series Analysis. Princeton University Press.

Appendix B. Estimation Results on Financial Performance

Table B.1 displays the descriptive analysis of the financial-performance variables for generic, focal-brand, and competing-brand keywords. The results in Table B.1 suggest that focal-brand and competing-brand keywords generate significantly more orders and more revenue than generic keywords. Total expenses for competing-brand keywords are lower than those for generic keywords. Finally, the profit of focal-brand and competing-brand keywords is significantly higher than that of generic keywords.

Table B.1 Descriptive Analysis of Financial Performance

		Mean	F value	p value
Order	Generic	.01	1960	.004
	Competing	.06		
	Focal	1.63		
Total revenue	Generic	2.68	628.222	.001
	Competing	4.22		
	Focal	300.72		
Total expenses	Generic	1.61	369.261	.000
	Competing	0.93		
	Focal	6.11		
Profit	Generic	1.07	616.035	.000
	Competing	2.23		
	Focal	294.61		

Model Specification and Estimation Approach on Financial Performance

We first model Order. Because Order is continuous and left censored at zero, we use a type I Tobit regression model. The Tobit model applies to analyzing a censored variable, where a proportion n_0 of the observations has $\mathrm{Order_i} = 0$ and a proportion n_1 observations has $\mathrm{Order_i} > 0$. Daily orders generated by keywords is a censored variable, because some keywords do not generate any orders on certain days. OLS regression will be inconsistent in this situation, because it uses only those n_1 observations with $\mathrm{Order_i} > 0$.

Thus, we assume that a continuous and latent variable $\mathrm{Order_i^*}$ generates the observed values of the daily Order of each keyword, $\mathrm{Order_i}$,

$$\text{Order}_i = \begin{cases} \text{Order}_i^*, & \text{if Order}_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

Order_i^* is further specified by a linear function of the independent variables, as follows:

$$\text{Order}_i^* = \alpha_1 + \alpha_2\text{Focal}_i + \alpha_3\text{Competing}_i + \alpha_4\text{Rank}_i + \alpha_5\text{Rank}_i \times \text{Focal}_i + \alpha_6\text{Rank}_i \times \text{Competing}_i + \alpha_7\text{Length}_i + \alpha_8\text{Length}_i \times \text{Focal}_i + \alpha_9\text{Length}_i \times \text{Competing}_i + \alpha_{10}\text{Broad}_i + \alpha_{11}\text{Phrase}_i + \alpha_{12}\text{Broad}_i \times \text{Focal}_i + \alpha_{13}\text{Phrase}_i \times \text{Focal}_i + \alpha_{14}\text{Broad}_i \times \text{Competing}_i + \alpha_{15}\text{Phrase}_i \times \text{Competing}_i + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, \delta^2)$.

Then, the probability of Order_i^* being unobserved is

$$\Pr(\text{Order}_i^* \leq 0) = \Phi\left(\frac{-X_i\alpha}{\delta}\right),$$

where X_i stands for the above independent variables and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

Thus, we can identify the parameters by maximizing the likelihood function:

$$\mathcal{L}(\alpha, \delta|\text{Data}) = \prod_{i=1}^{N} f(\text{Order}_i)$$

$$= \prod_{i=1}^{N} \left[\frac{1}{\sqrt{2\pi\delta^2}} \exp\left(-\frac{1}{2\delta^2}(\text{Order}_i - X_i\alpha)^2\right)\right]^{I(\text{Order}_i)} \left[\Phi\left(\frac{-X_i\alpha}{\delta}\right)\right]^{1-I(\text{Order}_i)}$$

Finally, Profit and ROI are specified as the following linear model:

$$\text{Profit}_i = \beta_1 + \beta_2\text{Focal}_i + \beta_3\text{Competing}_i + \beta_4\text{Rank}_i + \beta_5\text{Rank}_i \times \text{Focal}_i + \beta_6\text{Rank}_i \times \text{Competing}_i + \beta_7\text{Length}_i + \beta_8\text{Length}_i \times \text{Focal}_i + \beta_9\text{Length}_i \times \text{Competing}_i + \beta_{10}\text{Broad}_i + \beta_{11}\text{Phrase}_i + \beta_{12}\text{Broad}_i \times \text{Focal}_i + \beta_{13}\text{Phrase}_i \times \text{Focal}_i + \beta_{14}\text{Broad}_i \times \text{Competing}_i + \beta_{15}\text{Phrase}_i \times \text{Competing}_i + \upsilon_i$$

$$\text{ROI}_i = \gamma_1 + \gamma_2\text{Focal}_i + \gamma_3\text{Competing}_i + \gamma_4\text{Rank}_i + \gamma_5\text{Rank}_i \times \text{Focal}_i + \gamma_6\text{Rank}_i \times \text{Competing}_i + \gamma_7\text{Length}_i + \gamma_8\text{Length}_i \times \text{Focal}_i + \gamma_9\text{Length}_i \times \text{Competing}_i + \gamma_{10}\text{Broad}_i + \gamma_{11}\text{Phrase}_i + \gamma_{12}\text{Broad}_i \times \text{Focal}_i + \gamma_{13}\text{Phrase}_i \times \text{Focal}_i + \gamma_{14}\text{Broad}_i \times \text{Competing}_i + \gamma_{15}\text{Phrase}_i \times \text{Competing}_i + \varpi_i$$

and the estimation approach is straightforward by employing OLS.