SHOW ME THE INCENTIVES: A DYNAMIC STRUCTURAL MODEL OF EMPLOYEE BLOGGING BEHAVIOR

Yan Huang  
*Carnegie Mellon University*, yanhuang@cmu.edu

Param Vir Singh  
*Carnegie Mellon University*, psidhu@cmu.edu

Anindya Ghose  
*New York University*, aghose@stern.nyu.edu

Follow this and additional works at: [http://aisel.aisnet.org/icis2010_submissions](http://aisel.aisnet.org/icis2010_submissions)

**Recommended Citation**

[http://aisel.aisnet.org/icis2010_submissions/200](http://aisel.aisnet.org/icis2010_submissions/200)
SHOW ME THE INCENTIVES: A DYNAMIC STRUCTURAL MODEL OF EMPLOYEE BLOGGING BEHAVIOR

Completed Research Paper

Yan Huang
Heinz College
Carnegie Mellon University
yanhuang@cmu.edu

Param Vir Singh
Tepper School of Business
Carnegie Mellon University
psidhu@cmu.edu

Anindya Ghose
Stern School of Business
New York University
aghose@stern.nyu.edu

Abstract

Many firms believe that enterprise blogging forums can be used to help build the structured platform required for an environment that supports emergent innovation. While some employees tend to be ‘consumers’ of content created by others, some others contribute by acting as ‘creators’. In this paper, we build and estimate a dynamic structural model towards understanding the mechanisms that incentivize users to contribute to blog forums that are consumed by employees across the organization. We find strong evidence of competitive dynamics in our enterprise wide data setting. Our results demonstrate why employees contribute to blog forums. We find that employees derive higher utility from readership of their work related posts than their leisure-related posts. Employees compete with their peers to attract more readerships for their posts. Further, results indicate positive spillover effect from readership of leisure posts to work posts of a employee. We also find that knowledge-based benefits are higher for work related knowledge than leisure related knowledge. Our results suggest that enterprises would benefit more from feedback systems that provide a picture of how knowledge workers in the organization are interacting with the tools are made available to them. We discuss implications for implementing feedback systems that quantify the reputation of the content creator and incentivize employees to engage in this practice.

Keywords: Blog(s), Blogging community, User-generated-content, Web 2.0, Structural modeling, Econometric analyses, Economics of information systems, Enterprise 2.0, Utility

Introduction

Over the last few years, blogs have become one of the most prominent web 2.0 technologies. Blog search engine Technoratti reports that there were more than 113 million blogs by April 2010, with 7.5 million of them active. Increasingly several leading organizations are encouraging their employees to blog (Aggarwal et al. 2010; Lee et al. 2006; Singh et al. 2010a). Many firms believe that enterprise blogging forums can be used to help build the structured platform required for an environment that supports emergent innovation. Prominent among them are Sun Microsystems, Microsoft, Infosys, Cognizant, Google, etc. Thus, the success of corporate and industry blogs that actually provide updated and useful information has increased interest in blogs as a communication and discussion mechanism within the enterprise.
This drive by corporations to encourage employees to blog is because employee blogs have emerged as new sources of knowledge sharing within the enterprise (Huh et al. 2007; Lee et al. 2006; Singh et al. 2010a; Yardi et al. 2009). However, employees’ motivation to blog may not always be in alignment with the firm's objective. Employee motivations determine whether they blog or not, what kind of information they share through their blogs, and what kind of blogs they read in turn. These activities have implications for their reputation within the firm and could potentially influence their career path and internal progress. Because of the increasing pervasiveness of this phenomenon, it is becoming increasingly important to understand the employee motivations to blog in an enterprise setting. Such an understanding would help devise strategies for a firm to influence employee blogging behavior.

In this paper, we aim to present an empirical framework to analyze dynamics in enterprise blogging using a unique individual-level dataset that maps the dual role of blog posting and reading behavior by employees. Our proposed model considers the employee motivations for reading and writing posts within an enterprise setting. The model is flexible enough to handle the reputational concerns, knowledge seeking concerns, and the tradeoff between sharing pure entertainment-related knowledge (leisure) and work-related knowledge along with the time constraints. The model explicitly incorporates dynamics induced by these aspects of the employee behavior. We apply the model to a rich dataset that comprises the complete details of blog posting, blog reading, post content, and employee demographics for a large dataset of approximately 2400 employees over a period of one year at a Fortune 1000 IT services and consulting firm.

We classify information in blogs into two types: work-related and non-work (leisure) related. In our model, employees are forward-looking and they make their decisions to maximize their long term utility instead of being myopic. Employees compete with their peers to attract more readerships for their posts. Hence, their actions are their best response to the peers' actions. We then formulate an employee’s decision on when to write or read, what to write or read (in terms of work-related or leisure-related) as a dynamic competition game in the tradition of structural dynamic competition games such as Bajari et al. (2007); Pakes and McGuire (1994); Pakes and McGuire (2001), Benkard (2004) and Ericson and Pakes (1995). Following Erickson and Pakes (1995) we focus on Markov Perfect Equilibrium as a solution concept.

There are two main components of the employee's utility function, which forms the core of this paper. The first component of an employee's utility function is reputation. Individuals gain reputational benefits from writing posts in an enterprise setting. Blog writing allows individuals to express their expertise to a broad audience at a low cost. Hence, individuals may receive opinion leadership status, self-satisfaction, or indirect economic incentives, such as promotion, salary hike, etc, from writing posts (Aggarwal et al. 2010; Kavanaugh et al. 2006). The reputation incentive is proportional to the readership of an employees' blog. This is important because if an employee writes a post and no one reads it then he does not receive any reputational benefit from that investment. Further, the work-related and leisure-related posts may provide different kinds and levels of incentives. Individuals also derive utility from the knowledge they can acquire from reading posts (Huh et al. 2007; Singh et al. 2010a; Yardi et al. 2009). This knowledge base forms the second component. We allow the knowledge-based utility derived from the work and leisure-related posts to be different. The work-related knowledge increase may help them in becoming more productive, be more informed about new ideas or technologies, open up new opportunities for collaborations, etc. (Yardi et al. 2009, Huh et al. 2007). The leisure-related information feeds to an employee's interests, allows him to relax, and updates him about current affairs, sports activities, cultural information, etc Singh et al. (2010a).

Our results show that employees are forward looking. While the instantaneous benefit of blogging is very small, it has a long term significant effect. It is only in the long term that the benefits of blogging outweigh its cost. Employees derive higher utility from readership of their work-related posts than their leisure-related posts. We find intense competition among employees in regards to attracting readership for their posts. While readership of leisure posts provides very little direct utility employees still post a significant amount of these posts as there is a significant spillover effect on the readership of work posts from leisure posts. Further, employees derive knowledge based utility from reading both work and leisure related posts. Results indicate that it is harder for employees to absorb work related knowledge through blogs than leisure related knowledge. Further, in comparison to leisure posts, employees need to read work posts more often to keep up their work knowledge levels.

Our study aims to makes a number of contributions. First, it provides insights into a very important but confusing phenomenon of why individuals incur the cost of reading and writing blogs when there appear to be no explicit monetary incentives for doing so. Second, we provide a structural framework to analyze employee blogging activities. An advantage of structural model is that it explains the reasons for user behavior rather than fitting the data as a reduced model would do. Our econometric framework captures theoretical arguments of psychological,
economic, and social motivations for participating in these activities. Third, while prior studies have investigated why individuals blog, the results from those studies are based on surveys/questionnaires which can be affected by self-reporting bias. In contrast, our study uses actual micro-level blogging activity data from a large enterprise-wide setting to shed light on why individuals blog and model employee behavior accordingly.

**Literature Review**

In the last few years, the research on the different aspects of blogs has exploded. Recent studies have focused on understanding why people blog. Nardi et al. (2004) find that bloggers are driven to document their lives, provide commentary and opinions, express deeply felt emotions, and articulate ideas through writing. Miura and Yamashita (2007) surveyed a large number of bloggers to understand why blogger continue writing posts. They found that bloggers who are satisfied with the benefits their posts provide them within terms of self-expression, and relationship with others are more inclined to continue blogging. Cummings et al. (2002) and Mishne and Glance (2006) study why bloggers cite others posts.

Lin et al. (2006) provide a method to detect communities in blogosphere. Nakajima et al. (2005) develop methods to detect influential bloggers. Adar and Adamic (2005), and Todeva and Keskinova (2009) study large scale blog datasets and shed light on how information diffuses in a blogosphere. Recently, there has been an impetus to study blog reading behavior (Adamic and Glance 2005; Furukawa et al. 2007). These studies explain why individuals cite and comment on other posts, and join blog-rolls. They focus on readers who are also bloggers.

Since the adoptions of blogs within enterprises, some studies have begun to explore aspects of employee blogs. Aggarwal et al. (2010) study how negative posts by its employees can actually benefit a firm. They find that small number of negative posts are most beneficial to a firm. They argue that negative posts bring in more readers who are also exposed to the larger fraction of positive posts on the blog. Huh et al. (2007) conducted semi-structured interviews with fourteen active employee bloggers to investigate the role of blogging and its effects on work processes. They reveal that blogs facilitates access to tacit knowledge and resources vetted by experts, and, most importantly, contribute to the emergence of collaboration across a broad range of communities within the enterprise. Yardi et al. (2009) studied a large internal corporate blogging community using log files and interviews and found that employees expected to receive attention when they contributed to blogs, but these expectations often went unmet. Employees expected returns from their blogging activities and they expressed frustration if they invested time and received little or no perceived return on investment. Singh et al. (2010a) study blog reading dynamics of employees within a large firm. The find that most of the employees' time is devoted to reading and writing leisure-related posts. They also find that employees typically focus on one or two categories for some time and then switch to an exploratory phase where they read small amount in a number of categories after which they again switch to a reading heavily in a few categories.

Finally, this paper is also related to the emerging literature on the economic and social implications of user-generated content based on Web 2.0 technologies. Existing research focuses on the various factors that motivate users to generate content, especially in the digital media environment. Nov (2007) surveys Wikipedians and finds that fun and ideology are the major motives for UGC in a collaborative environment. A related stream of work has examined the economics of UGC in online communities using different media such as PC and mobile by modeling their impact on inter-temporal dependencies in user behavior Ghose and Han (2010a) as well as modeling the presence of dynamic learning in such two-sided forums Ghose and Han (2010b). From a social standpoint, the impact of blogs on outcome of important events such as political elections has been studied by Adamic and Glance (2005), Drezner and Farrell (2004), and Farrell et al. (2008).

**Dynamic Game Model**

**Model Setup**

Employees $i=1,...,I$ decide about blogging decisions on a periodic (e.g. weekly) basis for $t=1,...,T$. In the enterprise blogging there are two types of posts $j={w, l}$ where $w$ represents work-related posts and $l$ represents leisure-related posts. At each period, an employee decides whether to read (write) a type $j$ post. Let $a_t \in A_t$ denote consumer $i$'s actions at time $t$ and let $a_t = (a_{1t},...,a_{It})$ denote the set of time $t$ actions. Hence, $a_{it} = (d_{itwp},d_{itlp},d_{itwr},d_{itr})$, where $d_{itjp(r)}$ is an indicator variable which equals 1 if employee $i$ posted (read) a type $j$ post at time $t$.
We assume that the employee's per period utility function at time $t$ comprises of utility from reputation, knowledge, outside good, and an unobserved private shock. We combine all other factors that affect an individual’s utility function into outside goods. All these factors are explained and operationalized below.

**Reputation**

Constant et al. (1994) points out that sharing expertise may produce significant personal benefit in terms of increase in personal identification with workers and organizations. These benefits can be applied in blog settings, as intuitively, expected reputation gain is one of the most important motivations for individuals to write blogs. Employees derive reputation benefits from the posts they write. When an employee writes a post it indicates her expertise in an area. The employee derives reputational benefit consistent with her readership. In the utility function, the reputation-based utility is incorporated as $\omega_{it}$ where

$$\omega_{it} = \theta_1 R_{itwp} + \theta_2 R_{itlp}.$$  

Here, $R_{it/jp}$ is the number of readers for type $j$ posts written by employee $i$ in period $t$. We allow for the readership of the two types of posts to enter separately into the model as they may lead to different kinds and levels of incentives. Work-related posts may express a reader's expertise in work-related knowledge, which may help an employee derive indirect/direct economic incentives within the enterprise. The leisure-related posts may benefit the individual getting a fan following and becoming more popular among the employee who read her posts.

**Knowledge**

Blogs facilitate access to tacit knowledge and resources vetted by experts Huh et al. (2007). The primary reason why corporations allow their employees to participate in blogging activities during their work hours is because the employee blogs act as a new source of work relevant knowledge sharing within the enterprise (Huh et al. 2007; Lee et al. 2006; Singh et al. 2010a; Yardi et al. 2009). In the knowledge economy, knowledge is a key resource both for the firm and for employees. It makes a knowledge worker more productive leading to associated benefits (Singh et al. 2010c). Employees can acquire knowledge by reading other's posts. When employees read others posts they become more aware of their (blogger) expertise, which opens up new opportunities for collaborations. As in the case of reputation the two types of posts affect the type and level of incentives differently. Hence, we enter them separately. In the utility function, the knowledge-based utility is captured by $\tau_{ijt}$ where

$$\tau_{ijt} = \theta_3 K_{itw} + \theta_4 K_{itl}.$$  

Here $K_{it/j}$ is the knowledge of type $j$ for a employee $i$ at the end of time period $t$.

**Budget Constraint**

Every week an employee has 24X7 hours ($y_{it}$) to do all activities. Let $t_{rf}$ be the cost of identifying and reading one post of type $j$ ($j = w, l$) and $t_{pj}$ be the cost of developing and writing one post of type $j$. Then we have the following budget constraint:

$$y_{it} = \sum_j t_{rf} d_{it/jr} + \sum_j t_{pj} d_{it/jp} + t_0 O_{it}.$$  

Here, $O_{it}$ is the outside good consumption and $t_0$ is the associated coefficients that captures the per unit time cost of consuming outside good. This budget constraint allows us to capture the tradeoff that an individual would consider while deciding time to allocate to pure leisure or work activities. Let us define $\theta_5 = t_{rw} / t_0; \theta_6 = t_{rl} / t_0; \theta_7 = t_{pw} / t_0; \theta_8 = t_{pl} / t_0$. Then the budget constraint can be written as:

$$\frac{y_{it}}{t_0} = \theta_5 d_{itwr} + \theta_6 d_{itlw} + \theta_7 d_{itwp} + \theta_8 d_{itlp} + O_{it}.$$
Utility Function

An employee's utility at time $t$ is given by $U_{it}$ as:

$$U_{it} = \omega_{it} (\theta_1, \theta_2, R_{twp}, R_{tcp}) + \tau_{ijt} (\theta_3, \theta_4, K_{itw}, K_{ltt}) + y_{it} + \gamma_{it}.$$ 

$\gamma_{it}$ is the random shock to the utility that may affect an employee's decisions. We assume that the error $\gamma_{it}$ satisfies $\gamma_{it} = \sum_{jq} \gamma_{ijtq} x d_{ijtq}$, and $\gamma_{ijtq}$ has an i.i.d extreme value type one distribution. Note that the first component of the utility function is a function of both the individual’s state and the peers’ states.

The utility function can be further written as:

$$U_{it} = \theta_1 R_{twp} + \theta_2 R_{tcp} + \theta_3 K_{itw} + \theta_4 K_{ltt} + 0_{it} + y_{it}.$$ 

Substituting the budget condition into the utility function gives:

$$U_{it} = \theta_1 R_{twp} + \theta_2 R_{tcp} + \theta_3 K_{itw} + \theta_4 K_{ltt} + \frac{y_{it}}{t_o} - \theta_5 d_{itw} - \theta_6 d_{ltt} - \theta_7 d_{twp} - \theta_8 d_{tcp} + y_{it}.$$ 

Since $y_{it}/t_o$ affects all choices in the same way, we drop it from the utility function.

State Variables

There are four state variables corresponding to each individual in our model. The first and the second state variables are the readership variables $R_{twp}$ and $R_{tcp}$. The readership variables evolve as follows:

$$R_{itjp} = \delta R_{it-1jp} + \eta_{itjp}.$$ 

Here, $\eta_{itjp}$ is the number of people who read $i$'s type $j$ post in period $t$. In addition, $\delta$ is a depreciation factor which is set at 0.9. The value of 0.9 is chosen arbitrarily. Our results are robust to alternate values of the depreciation factor. At the beginning of period $t$ an employee will not know $\eta_{itjp}$. However, given that the behavior observed in the data is an equilibrium behavior, and the rational expectation equilibrium would suggest that the employee's ex ante expectation of readership would match ex post realization of readership. To operationalize this, we assume that the individual has a belief about the mean and variance of the readership he may receive for his/her new post. In equilibrium this mean and variance would match the true mean and variance. We estimate the mean of the expected new readership semi parametrically as a function of individuals', his/her peers' states, and actions (Bajari et al. 2009).

The third and fourth state variables are the cumulative knowledge of employee $i$, $K_{itw}$ and $K_{ltt}$. The knowledge of an individual is unobserved to us. Hence, we follow Arcidiacono and Miller (2006) and employ a hidden Markov model (HMM) framework to identify the knowledge state and its evolution. Following Singh et al. (2010b), we consider knowledge to be discrete state and enforce that an individual’s reading of type $j$ post increases her probability of moving to a higher state. Using Singh et al. (2010b), we are able to retrieve back the most probable knowledge state for each individual in each time period. Due to space limitations, we do not provide the details of the HMM and semi parametric estimations here.

We define $s_a$ as the set of the state variables for an employee. Then $s_{it} = (R_{itwp}, R_{tcp}, K_{itw}, K_{ltt})$ is value of the state variables for an employee $i$ at time $t$. We further define $s_{-it} = (R_{-itwp}, R_{-tcp}, K_{-itw}, K_{-ltt})$ as the set of state variables of $i$’s peers. Then the strategy profile for $i$ is defined as $s = (s_{it}, s_{-it})$. One important point about strategy profile is that ideally, $s$ represents the states of all employees. In our case, however, there are around 2400 employees. Tracking every employee’s state would make our model unmanageable. To deal with this problem, we introduce a simplifying but more realistic assumption that employees make their decision according to their own state and the average of all employees Aguirregabiria and Ho (2010). This is a reasonable assumption because in reality, it is infeasible for employees to track all other employees’ states and make decisions accordingly. Instead, employees may only want to get a general idea about what the other employees’ states moments are and make decision based on his/her own state and the moments. In other words, strategy is a function of employees’ own states and the moments of reputation states of the whole competing group.
**Sequence of Events**

The specific sequence of events in our model is as follows:

1. An employee observes his state \( s_{it} \).
2. An employee receives random shocks \( y_{it} \) for the reading and writing decisions.
3. An employee makes an expectation of the readership \( r_{it,jp} \) he may get in the current period if he were to write a post. Note that the employee probability of moving up a knowledge state is just a function of his/her current knowledge state and reading action. The employee has perfect information about the knowledge state transition probabilities.
4. An employee makes decisions with regard to what to write or read in the current period.
5. An employee receives utility from his decision.
6. An employee’s state evolves because of his decision.

**Long Term Utility Function**

We model the employee's writing and reading decisions as a dynamic optimization problem. The employee's tasks are to decide when and whether to read work post, whether to read leisure post, whether to write on work, whether to write on leisure to maximize the sum of the discounted expected future utility over the infinite horizon.

\[
\max_{d_{itwp}, d_{itip}, d_{iterr}, d_{itir}} E_t \left( \sum_{t=0}^{\infty} \beta^{t-t} U_{it} \right)
\]

The variable \( \beta \) is the common discount factor. The operator \( E_t[\cdot] \) denotes the conditional expectation operator given the employee's information at time \( t \). There are two components of the model that need to be emphasized. The employee in our model maximizes his life time utility, which makes the model dynamic. At the same time, the utility of an employee is a function of the decisions made by his/her peers (through \( r_{it,jp} \)) making it a multi-agent dynamic game.

**Equilibrium Concept**

Following Ericson and Pakes (1995) we focus on Markov Perfect Equilibrium (MPE) as a solution concept. We assume that the behavior is consistent with MPE. In an MPE, each employee's behavior depends only on the current states and his current private shock. Formally, a Markov strategy for an employee \( i \) is a function \( \sigma_i : S \times I_t \rightarrow A_i \). A profile of Markov strategies is a vector, \( \sigma = (\sigma_1, ..., \sigma_t) \), where \( \sigma : S \times I_1 \times ... \times I_k \rightarrow A \). If behavior is driven by a Markov strategy profile \( \sigma \), employee \( i \)'s expected utility given state \( s \) can be written recursively as Bellman Equation:

\[
V_i(s; \sigma) = E_{it} \left[ U_i(\sigma(s,y), s, y_{it}) + \beta \sum_{s'} V_i(s' ; \sigma) dP(s' | \sigma(s,y), s) \right] s.
\]

Here, \( V_i \) is a value function which reflects expected value for employee \( i \) at the beginning of a period before private shocks are realized. Following the literature, a profile \( \sigma \) is Markov perfect equilibrium if, given the opponent profile \( \sigma_{-i} \), each employee \( i \) prefers its strategy \( \sigma_i \) to all alternate strategies \( \sigma_i' \). That is, for \( \sigma \) to be MPE

\[
V_i(s; \sigma, \sigma_{-i}) \geq V_i(s; \sigma_i', \sigma_{-i}).
\]

**Empirical Estimation**

The model parameters that need to be estimated are: \( \rho = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8) \). Ideally, one would like to estimate the model with a backward induction process. One would also need to consider the complexity raised by the presence of dynamic competition. As shown by research on dynamic competition, by Ericson and Pakes (1995), Pakes and McGuire (2001), and Benkard (2004), computing equilibrium for even relatively simple models is all but prohibitive. Hence, estimating the dynamic parameters is all but prohibitive.
To estimate the structural model, we follow the two-step estimation procedure suggested by Bajari et al. (2007) (hereafter BBL) for this scenario. The BBL estimator builds on the work by Rust (1987) which proposes an estimation strategy for single agent dynamic discrete choice models. There are several other estimation procedures suggested by for example Pakes et al. (2007), Pesendorfer et al. (2003), and Aguirregabiria and Mira (2007). However, all of them apply only to the case of discrete state dynamic games. However, the BBL estimator applies to both the discrete and continuous state cases. We combine the BBL estimator with the Bajari et al. (2009) and Arcidiacono and Miller (2006) to estimate the model.

**Two step approach for estimation**

In the first step we estimate the policy functions \( (\sigma_t; S \times I_t \rightarrow A_t) \), the state transition probabilities \( P: A \times S \rightarrow \Delta S \), and the Value functions. Our choice variables are discrete. And the utility function defined above implies additive separability and conditional independence between \( I_t \) and \( Y_t \). Define the choice-specific value function as

\[
v_t(a_i, s_i) = E_{Y_t}[P_t(a_i, \sigma_s(s_{i-1}, v_{-i}), s_i) + \beta \int V_t(s'_i, \sigma_v(s_{i-1}, v_{-i}), s_i) ds_i]
\]

With this notation, action \( a_i \) is employee \( i \)'s optimal choice when

\[
v_t(a_i, s_i) + \gamma_i(a_i) \geq v_t(a^*_i, s_i) + \gamma_i(a^*_i), \forall a^*_i \in A_i
\]

Hotz and Miller (1993) showed how to recover the choice-specific value functions by inverting the observed choice probabilities at each state:

\[
v_t(a^*_i, s_i) - v_t(a_i, s_i) = \ln(Pr(a^*_i|s_i)) - \ln(Pr(a_i|s_i))
\]

Given this, we only need to estimate the distribution of actions at each state, i.e. CCP, from the data. Note that we do not observe knowledge states. Hence, as discussed earlier, we follow Arcidiacono and Miller (2006) and employ an HMM to identify the knowledge states, the Markovian transition probabilities of the knowledge states as a function of actions and the most probable knowledge state for each individual every time point. Once we obtain the knowledge state of each individual, we treat it as an observed state. However, several of state variables are continuous (reputation states and the average states of peers). Therefore, \( Pr(a_i|s_i) \) will be a function, rather than discrete values. Following Bajari et al. (2009), we use sieve logit method to estimate \( \eta_{qit}^q \) with a second degree orthogonal polynomial as basis.

The estimation strategy suggested by BBL is inspired by Hotz and Miller (1993). The estimation process uses forward simulation to estimate an employee's value functions for a given strategy profile (including the equilibrium profile) given an estimate of state transition probabilities. Let \( V_t(s; \sigma; \rho) \) denote the value function of employee \( i \) at state \( s \), assuming \( i \) follows Markov strategy \( \sigma_t \) and all its peers follow \( \sigma_{-i} \). Then

\[
V_t(s; \sigma; \rho) = E \left[ \sum_{t=0}^{\infty} \beta^t U_t(\sigma(s_t, \gamma_t), s_t, \gamma_t; \rho) \right] \bigg| s_0 = s; \rho
\]

Given the first stage state transition probability estimates, we can use simulation to estimate the value function \( V_t(s; \sigma; \rho) \) for any strategy profile \( \sigma \) and parameter vector \( \rho \). A simulated path of play is obtained as:

**Step 1**: Starting at state \( s_0 = s \), draw private shocks \( \gamma_{i0} \) from \( G_t(\cdot | s_0; \rho) \) for every employee \( i \).

**Step 2**: Calculate specified action \( a_{i0} = \sigma_t(s_0, \gamma_{i0}) \) for each employee \( i \) and the resulting utility \( U_{i0}(a_{i0}, s_0, \gamma_{i0}; \rho) \).

**Step 3**: Determine the new state \( s_{i1} \) according to the state transition function.

**Step 4**: Repeat the steps 1-3 for \( T \) period.

If the policy profile used in this step is the policy profile estimated in the first stage then it is the resultant value over a number of simulated paths is an estimate of the payoff \( (V_t(s; \sigma; \sigma_{-i}; \rho)) \) from playing \( \sigma_t \) in response to all peers playing \( \sigma_{-i} \).

In the second step, we use the estimates from first stage combined with the equilibrium conditions of the model to estimate the underlying structural parameters.
Let \( z \in Z \) index the equilibrium conditions, so that each \( z \) denotes a particular \((i, s; \sigma')\) combination. Let us further define:

\[
g(z; \rho, \kappa) = V_l(s; \sigma, \sigma_{-i}; \rho, \kappa) - V_l(s; \sigma', \sigma_{-i}; \rho, \kappa).
\]

Here, \( \kappa \) reflects that \( \sigma \) and state transitions are parameterized by \( \kappa \). The equilibrium condition is satisfied at \( \rho, z \) if \( g(z; \rho, \kappa) \geq 0 \). This estimated through simple simulated minimum distance estimator.

**Identification and Normalization**

There are several identification issues that need to be addressed before the model can be consistently estimated. First, the distribution of private shocks is assumed to be known for identification. We assumed that the private shocks are extreme value type 1 distributed. Second, in HMM the probability of knowledge increase should be higher for an individual as a result of reading than otherwise. We enforce this by ensuring that reading has a non-negative impact on knowledge increase. We cannot identify \( t_0 \) and he cost specific parameters together. Hence, we normalize \( t_0 = 1 \).

Consistency and asymptotic normality of estimates is established as sample size, number of simulated paths, and number of alternatives, go to infinity at appropriate rates. Efficiency depends on chosen alternatives. We perturb the true policy function enough to generate enough variation for efficiency.

We do not know the initial values of the state variables, which raise the well known 'initial conditions' problem. The first observation in our sample may not be the true initial outcome of an employee's blog content generation and reading behavior. If one does not control for initial choice history, the implicit assumption is that every employee has the same beliefs across both kinds of activity (read and write) and across both kinds of blogs (work and leisure). However, it is possible that a employee that has engaged in an activity multiple times in the past would have more informed priors than another user who has engaged very little in that activity. Hence, one needs to account for the heterogeneity of priors in the sample. We follow an approach that is similar in spirit to that used in Erdem et al. (2008). We use a part of our data (4 months) as a pre-estimation sample to estimate the distribution of state variables.

**Data Description**

Our research setting is a large, Fortune 1000 information technology services, business process outsourcing, and consulting firm. Fortune named this firm one of the fastest growing companies in 2009 (Fortune 2009). Its annual revenues in year 2009 were a few billion dollars. It is a US-based firm with significant presence and operations in several other countries across multiple continents: Europe, Asia, and Americas being the major areas. To influence more knowledge and information sharing across as well as within locations, the firm has undertaken several measures. Prominent among these measures is the use of Web 2.0 technologies within the enterprise.

This firm adopted the use of employee blogs in late 2006. These blogs are hosted on an internal platform and are not accessible by people outside the organization. Every employee is allowed to host his/her own blog on this platform. These blogs are accessible to all the employees of the firm across the entire hierarchy. The identity of the blogger is also revealed on the blog. Bloggers classify their posts into one of the 25 categories (for example, software testing, movies, history, knowledge management, senior management, etc). To be able to measure the knowledge sharing aspect, the firm tracks who (which employee) reads which blog and at what time. Since the blogs are only internally accessible, the firm did not impose any restrictions on the kind of posts that can be written by the employees. To analyze the type of content that is being shared in the internal blogosphere, the firm broadly classifies the blog article categories into two topics: Work-related \((w)\) and Leisure-related \((l)\). Table 1 presents the sub-categories that constitute each topic.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Leisure Related</th>
<th>Work Related</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sub Categories</strong></td>
<td>Fun; Movies-TV-Music; Sports; Puzzles; Chip-n-putt; Religion-Spiritual-Culture; History-</td>
<td>FLOSS; Technology; Testing; Domains; Corporate Functions; Knowledge Management; Project</td>
</tr>
</tbody>
</table>
For estimation purposes, we randomly selected 2396 employees of the firm. We collected data about their blogging activities for approximately 64 weeks. We have data on exact timestamps of blog reading and posting activities. For the purpose of estimation, we define a period as one week. This provides us data for 64 weeks.

High level descriptive statistics of our data are shown in Table 2. This set of employees wrote 41078 posts during the 64 week period. Out of these 14474 posts are work-related and 26604 posts are leisure related. Variable definitions along with associated parameters are shown in Table 3. The descriptive statistics for the key variables used to construct the model variables are presented in Table 4.

Table 2. Overall Sample Statistics

<table>
<thead>
<tr>
<th>Overall Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Employees</td>
</tr>
<tr>
<td>Period length</td>
</tr>
<tr>
<td>Number of periods</td>
</tr>
<tr>
<td>Total posts written</td>
</tr>
<tr>
<td>Work-related posts written</td>
</tr>
<tr>
<td>Leisure-related posts written</td>
</tr>
<tr>
<td>Work-related post reading</td>
</tr>
<tr>
<td>Leisure-related post reading</td>
</tr>
</tbody>
</table>

Table 3. Variable Meaning and Corresponding Parameters

<table>
<thead>
<tr>
<th>Variable</th>
<th>Corresponding Parameter</th>
<th>Variable Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{tw}/1000$</td>
<td>$\theta_1$</td>
<td>Scaled cumulative reputation (measured by depreciated past readership and current period readership) of work-related posts for employee $i$ in period $t$.</td>
</tr>
<tr>
<td>$R_{ltw}/1000$</td>
<td>$\theta_2$</td>
<td>Scaled cumulative reputation (measured through depreciated past readership and current period readership) of leisure-related posts for employee $i$ in period $t$.</td>
</tr>
<tr>
<td>$K_{tw}$</td>
<td>$\theta_3$</td>
<td>Work-related Knowledge state (measured in levels. The number of levels needs to be specify after HMM estimation)</td>
</tr>
<tr>
<td>$K_{lt}$</td>
<td>$\theta_4$</td>
<td>Leisure-related Knowledge state (measured in levels. The number of levels needs to be specify after HMM estimation)</td>
</tr>
<tr>
<td>$d_{tw}$</td>
<td>$\theta_5$</td>
<td>Average cost of reading work-related posts in terms of time</td>
</tr>
</tbody>
</table>
\(d_{itlr}\) \(\theta_6\) Average cost of reading leisure-related posts in terms of time  
\(d_{itwp}\) \(\theta_7\) Average cost of writing work-related posts in terms of time  
\(d_{itlp}\) \(\theta_8\) Average cost of writing leisure-related posts in terms of time

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R_{itlp})</td>
<td>16.28</td>
<td>91.34</td>
</tr>
<tr>
<td>(R_{itlw})</td>
<td>86.19</td>
<td>305.22</td>
</tr>
<tr>
<td>(K_{itw})</td>
<td>0.12</td>
<td>0.33</td>
</tr>
<tr>
<td>(K_{ittl})</td>
<td>0.16</td>
<td>0.37</td>
</tr>
<tr>
<td>(d_{itwr})</td>
<td>0.17</td>
<td>0.38</td>
</tr>
<tr>
<td>(d_{itlr})</td>
<td>0.24</td>
<td>0.43</td>
</tr>
<tr>
<td>(d_{itwp})</td>
<td>0.06</td>
<td>0.18</td>
</tr>
<tr>
<td>(d_{itlp})</td>
<td>0.10</td>
<td>0.26</td>
</tr>
</tbody>
</table>

### Results

Here we present the results for first and second stage separately.

#### First Stage Results

**Knowledge States**

We identify two distinct knowledge states for both work and leisure. We classify these states as high and low with respect to knowledge. Table 5 and 6 provide the transition probabilities of work and leisure knowledge states respectively.

**Table 5. Work Knowledge State Transition Probabilities**

<table>
<thead>
<tr>
<th>t (\rightarrow) t+1</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.99 (0.89)</td>
<td>0.01 (0.11)</td>
</tr>
<tr>
<td>High</td>
<td>0.63 (0.19)</td>
<td>0.37 (0.81)</td>
</tr>
</tbody>
</table>

The probabilities outside (inside) parenthesis indicate the probabilities with (no reading) reading.

**Table 6. Leisure Knowledge State Transition Probabilities**

<table>
<thead>
<tr>
<th>t (\rightarrow) t+1</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.88 (0.77)</td>
<td>0.12 (0.23)</td>
</tr>
<tr>
<td>High</td>
<td>0.52 (0.05)</td>
<td>0.48 (0.95)</td>
</tr>
</tbody>
</table>

The probabilities outside (inside) parenthesis indicate the probabilities with (no reading) reading.

Several interesting observations can be made from Table 5 and 6. It is hard to get to a high work knowledge state in comparison to leisure knowledge state from the corresponding low state. Further, once an individual is in high state it is easier for her to stay in the high leisure knowledge state than work knowledge. The matrixes indicate that it is much harder to acquire work knowledge than leisure knowledge from blog reading.

**Reputation States**

The reputation state is calculated as sum of discounted part cumulative readership and the new readership. While the past readership is a given at the beginning of the period, the new readership is estimated from a truncated normal.
distribution with a mean and variance. The mean and variance of this distribution are the function of individual’s state and peers’ states. We estimate the mean and variance semi parametrically from the data following Bajari et al. (2009). While space limitations prohibit us from reporting the full results here, we report some of the interesting insights here from this procedure. The results indicate a strong competition for readership among employee bloggers. Employees who have higher (lower) readership than their peers derive proportionately higher (lower) readership than peers. Similarly, as the peers’ readership increases an employee’s readership decreases. Further, there is a strong positive readership spillover effect from leisure posts to work posts. Given everything else equal, an individual who has high readership of his leisure posts would derive even higher readership of his work posts in comparison to an employee with low readership of leisure posts. Further, the readership of both work and leisure posts increase with average knowledge states. This indicates that when employees gain from reading posts they continue reading in future.

**Conditional Choice Probabilities**

As discussed earlier conditional choice probabilities are estimated semi parametrically. Space limitations do not allow us to report the full results here. The CCP highlight several interesting interdependencies among the state variables. In particular, there are several interesting interactions between work related activities and leisure related activities due to spillover effects. We discuss a few interesting results from the CCP here.

**Work Posting.** Results indicate that an employee is more likely to post if he has higher work readership and leisure readership than his peers. While an individual is less likely to post as the mean work readership of his peers increase, he is more likely to post if mean leisure readership of his peers increase. Further an individual is more likely to post if he and his peers have high work knowledge than otherwise. An individual is less likely to post if he and his peers have high leisure knowledge than otherwise.

**Leisure Posting.** Results indicate that an employee is more likely to post if he has higher work readership than his peers. An employee is also more likely to post leisure as his leisure readership increases only up to a point beyond which he is less likely to post as leisure readership increases. This suggests that beyond a certain point, employees are reluctant to be identified as top-ranked leisure bloggers while they perceive no disutility from being identified as top-ranked work bloggers. While an individual is less likely to post as the mean leisure readership of his peers increase, he is more likely to post if mean work readership of his peers increase. Further an individual is more likely to post if he and his peers have high leisure knowledge than otherwise. An individual is less likely to post if he has low work knowledge and his peers have high leisure knowledge than otherwise.

**Work Reading.** An employee is more likely to read if his and peers work knowledge and work readership is high. He is also more likely to read if his leisure knowledge and readership are high but peers leisure knowledge and reputation are low.

**Leisure Reading.** An employee is less likely to read if his work readership is low compared to peers and leisure readership is high compared to peers. While the employee is more likely to read if his work knowledge is high and his and his peers’ leisure knowledge is high, he is less likely to read if peer’s work knowledge is high.

**Second Stage Results**

The results for the second stage (structural parameters) are presented in Table 7. The results show that \( \theta_1 \) and \( \theta_2 \) are positive and significant. This indicates that employees gain positive utility from cumulative work-related readership and cumulative leisure-related readership. This set of results verifies that reputation gain provide employees incentives to write blogs. However, the magnitude of utility derived from work-related readership is almost 4 times of utility derived from leisure-related readership, indicating that in the enterprise blog environment, people appreciate work-related reputation more than leisure-related reputation.

From HMM, we know that for both work knowledge and leisure knowledge, there are only two state levels. We define the two levels high and low. The output of HMM also implies that employees have higher probability to stay
in high level at the end of a period if they read in the current period, and have lower probability to stay in high if they do not read. The knowledge states actually count for both past knowledge base and addition knowledge acquired in current period. Individuals derive knowledge-based utility from reading other’s posts as indicated by positive and significant $\theta_3$ and $\theta_4$.

### Table 7: Full Sample Structural Model Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimated values</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$</td>
<td>4.357***</td>
<td>0.254</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>1.157***</td>
<td>0.363</td>
</tr>
<tr>
<td>$\theta_3$</td>
<td>0.522***</td>
<td>0.027</td>
</tr>
<tr>
<td>$\theta_4$</td>
<td>0.018***</td>
<td>0.002</td>
</tr>
<tr>
<td>$\theta_5$</td>
<td>2.986***</td>
<td>0.169</td>
</tr>
<tr>
<td>$\theta_6$</td>
<td>0.726***</td>
<td>0.193</td>
</tr>
<tr>
<td>$\theta_7$</td>
<td>1.530***</td>
<td>0.054</td>
</tr>
<tr>
<td>$\theta_8$</td>
<td>0.834***</td>
<td>0.024</td>
</tr>
</tbody>
</table>

The parameters $\theta_5$ to $\theta_8$ are positive and significant. Note that in our utility function, we have a negative sign in front of these four parameters, hence, the parameters are expected to be positive. Also note that $\theta_5 - \theta_8$ are normalized costs, relative to the cost of consuming outside goods. Hence, the costs parameters can be compared with each other. Our estimates suggest that reading and writing work-related post is more time consuming than reading and writing leisure-related posts. One possible explanation could be that people who read or write work-related blog spend a lot of time digesting or creating the knowledge they learn from the blogs.

### Discussion and Conclusions

As enterprise blogs become an increasingly popular tool for generating and sharing information, this motivates the need to have a deeper understanding of users' content generation and usage behavior in organizational settings. While some employees tend to be 'consumers' of content created by others, especially their peers, some others contribute by acting as 'creators'. For a given user, content generation and usage may not be independent decision-making processes. Rather, they are likely to be inter-related processes. However, little is known about how content generation by users is related to their usage of such content, or vice-versa, and how these processes are associated with employees' organizational status within a firm. Moreover, this will vary based on the two prominent kinds of blogs that tend to exist within firms: work related and leisure-related blogs.

In an enterprise setting, success of a blog is even more setting as there may be potential career and promotion-related benefits associated with increased visibility of an employee's blog posting. Blog post generation is a costly task where an employee can adopt a forward-looking decision-making rule keeping in mind long-term objectives of enhancing his internal reputation and knowledge, given the opportunity costs of time associated with these activities. In the same vein, a given employee is also likely to observe how his peers behave and then take actions in order to maximize the reputation and knowledge-related gains in this setting after considering how current actions can influence the future decisions of peers, which in turn can affect the future utility of the focal employee. This makes the interactions between users strategic. One can thus envision that enterprise blogging is a ripe area for a dynamic competitive game between employees.

In this paper, we build and estimate a theory-based dynamic structural model towards understanding the mechanisms that drive users to contribute to blog forums that are consumed by their peers and other employees across the organization. We find strong evidence of such dynamics in our enterprise wide data setting. Our model recognizes not only the possibility that employee contribution decisions are inter-related across peers within the organization, but also that content generation (blog writing) and usage (blog reading) decisions are inter-temporally related. Because we explicitly model the process of inter-temporal user decision-making, our approach illustrates the micro-mechanisms of contribution and usage decisions in an enterprise blog setting, rather than merely testing whether interdependence exists using a reduced form approach.
So what can organizations do to spur knowledge creation through social media tools like blogs and wikis? Our results highlighting the effects of reputation, knowledge, and knowledge sharing suggest that enterprises would benefit more from feedback systems that provide a picture of how knowledge workers in the organization are interacting with the tools you make available to them. Better yet, these feedback systems should generate metrics that quantify the reputation of the content creator and incentivize people to engage in this practice. This can be done by adding some basic instrumentation to the knowledge sharing system. For example, they can make it simple to count things like blog posts made, comments made, documents contributed, documents consulted, and pointers shared and make these metrics salient and visible to all employees. Such metrics can thus help senior management evaluate junior employees on their knowledge sharing performance and use it as an additional tool in their annual evaluation. Furthermore, organizations can use such data to distill and identify patterns of content sharing practice that are worth emulating.

Our paper has some limitations, which could act as fruitful areas for future research. We have not explored the role of geographical location and physical proximity in our setting. It is plausible that distance from peers plays a role in the competitive dynamics within enterprises. We have also not examined the textual content of the blog postings. Future work can examine the sentiments and polarity of the blog content. Such information can be used by competitors to predict future readership, which will affect the decisions they take. Notwithstanding these limitations, we hope our paper, which provides the first known dynamic structural model of blogging, paves the way for future research in this important area.

References


