University of Pennsylvania
ScholarlyCommons

Finance Papers

Wharton Faculty Research

12-2015

A Structural Model of Employee Behavioral Dynamics in Enterprise Social Media

Yan Huang

Param Vir Singh

Anindya Ghose

Follow this and additional works at: http://repository.upenn.edu/fnce_papers

Part of the Economics Commons, and the Finance and Financial Management Commons

A Structural Model of Employee Behavioral Dynamics in Enterprise Social Media

Abstract

We develop and estimate a dynamic structural framework to analyze the social-media content creation and consumption behavior of employees within an enterprise. We focus, in particular, on employees' blogging behavior. The model incorporates two key features that are ubiquitous in blogging forums: users face (1) a trade-off between blog posting and blog reading; and (2) a trade-off between work-related and leisure-related content. We apply the model to a unique data set comprising the complete details of the blog posting and reading behavior of employees over a 15-month period at a Fortune 1000 IT services and consulting firm. Despite getting a higher utility from work-related blogging, employees nevertheless publish a significant number of leisure posts. This is partially because the creation of leisure posts has a significant positive spillover effect on the readership of work posts. Counterfactual experiments demonstrate that leisure-related blogging has positive spillovers for work-related blogging, and hence a policy of abolishing leisure-related content creation can inadvertently have adverse consequences on work-related content creation in an enterprise setting. When organizations restrict leisure blogging, the sharing of online work-related knowledge decreases and this in turn can also reduce employee performance rating. Overall, blogging within enterprises by employees during their work day can have positive long-term benefits for organizations.

Keywords

structural modeling, dynamics, enterprise social media, blog posting, blog reading, work-related content, leisure-related content

Disciplines

Business | Economics | Finance and Financial Management

A Structural Model of Employee Behavioral Dynamics in Enterprise Social Media

Yan Huang
yphuang@umich.edu
Ross School of Business
University of Michigan
& iLab Heinz College

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

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

Abstract

We develop and estimate a dynamic structural framework to analyze social-media content creation and consumption behavior of employees within an enterprise. We focus, in particular, on employees' blogging behavior. The model incorporates two key features that are ubiquitous in blogging forums: Users face 1) a trade-off between blog posting and blog reading; and 2) a trade-off between work-related and leisure-related content. We apply the model to a unique dataset that comprises the complete details of blog posting and reading behavior of 2389 employees over a 15-month period at a Fortune 1000 IT services and consulting firm. We find that the utility employees derive from work-related blogging is between 27% to 71% higher than what they derive from leisure-related blogging. However, employees still post a significant amount of leisure posts. This is partially because there is a significant spillover effect on the readership of work posts from the creation of leisure posts. In addition, we find that reading and writing work-related posts is more costly than reading and writing leisure-related posts, on average. We find evidence of strong competition among employees with regard to attracting readership for their posts. We conduct counterfactual experiments that provide insights into how different policies may affect employee behavior. We find that a policy of prohibiting leisure-related activities can hurt the knowledge sharing in an enterprise setting by reducing the work posting by 90% and work reading by 47%. By demonstrating that there are positive spillovers from leisure-related blogging to work-related blogging, our results suggest that a policy of abolishing leisure-related content creation can inadvertently have adverse consequences on work-related content creation in an enterprise setting.

Keywords: Structural modeling, Dynamics, Enterprise social media, Blog posting, Blog reading, Work-related content, Leisure-related content.

1. Introduction

Over the last two to three years, several firms have adopted various kinds of social-media technologies, such as blogs and wikis, for internal use. Increasingly, several leading organizations have systems in place to encourage their employees to blog (Lee et al., 2006; Aggarwal et al., 2011; Singh et al., 2010a). Prominent adopters are General Motors, IBM, HP, Microsoft, Infosys, Google, Charles Schwab, among others. Internally, organizational adoption of social media outlets such as enterprise blogs have been argued to improve access to knowledge experts, increase the rate of solving problems, increase the rate of new-product development, and reduce costs of internal communication (McAfee, 2006). The general thinking within such firms is that enterprise blogging forums can provide a structured platform that supports employee participation in brainstorming and idea sharing, thus deepening a company's pool of knowledge. Therefore, the use of blogs as a mechanism for knowledge creation and dissemination has increased their adoption in enterprises (Lee et al., 2006; Huh et al., 2007; Yardi et al., 2009).

However, employees' motivation to blog may not always be in alignment with a firm's objectives. Work- and leisure-related content often co-exist on the same page in enterprise blogging platforms.[1] Hence, some employees' blog posts may not be relevant to their work knowledge or professional expertise. Some firms believe that such behavior can undermine the very goal of enterprise blogging. Therefore, it is increasingly important to understand employees' blogging behavior in an enterprise social-media setting in order to gain insights about how firms' policy interventions can influence employee behavior. Typically, employees' propensity to create or consume content on a blog is constrained by the time available for such activities. Hence, their utility from blogging is determined by a trade-off between work- and leisure-related content. There is another trade-off that employees make: how much time to spend on content generation (i.e., posting blogs) vs. content consumption (i.e., reading blogs).

Employees can benefit from the readership they gain from blogging in an enterprise setting in a number of different ways. On the one hand, work-related blogging allows individuals to express their expertise to a broader audience at a relatively low cost and make a reputation for themselves. Once employees are identified as "experts" in certain domains, they may benefit from economic incentives, such as promotions and salary hikes (Kavanaugh et al., 2006; Aggarwal et al., 2011). Leisure-related blog posting, on the other hand, can help employees become popular among their peers and become "opinion leaders" on select topics. There are benefits to reading blogs, as well. An increase in work-related knowledge from reading blogs may help employees become more productive professionally, be more informed about new ideas, open up new opportunities for collaborations, and so on (Huh et al., 2007; Yardi et al., 2009). Leisure-related

---

[1]In this paper, by leisure-related blogs, we refer to all blog content that is not work-related. This includes topics such as travel, sports, entertainment, photography etc. More details are provided in Table 2.

information can feed employees' interests and allow them to relax (Singh et al., 2010a), thereby indirectly improving their subsequent productivity.

Readership gained from blogging is typically a measure of the quality of one's contribution over a period of time. An employee with a high work-related readership accumulated over time can still obtain readership-based reputational benefits even if he/she does not post in the current period. In contrast, an employee with a low work-related readership will find it much harder to build a high readership status based only on current-period posting. Since posting-related readership accrues based on employees' blogging contributions over a longer time horizon, the reputational benefits of blogging depend on the "cumulative readership" of an employee over time. Similarly, an employee's knowledge gain from reading blogs can lead to an improvement in productivity and performance not only in the current period, but also over a longer time horizon (McAfee, 2006)[2]. Hence, we theorize that employees in an enterprise setting are forward-looking, rather than myopic, when making their decisions about the trade-offs described above.

Though "cumulative work readership" and "cumulative leisure readership" provide employees with very different benefits, these two types of readership can have a strong interdependence that can be either positive or negative. Readers may follow particular bloggers, not specific topics or posts. Work or leisure posts by a given blogger appear on the same forum. Hence, employees with high levels of cumulative leisure readership can attract readers to their work-related posts, and vice versa. If this occurs, there can be positive spillover between work and cumulative leisure readership. However, it is also possible that employees with high cumulative leisure readership are seen as "unprofessional" in a work environment. If this case, readers would not expect their work-related posts to be of high quality; thus a high leisure readership can have a negative impact on an individual's work readership. Because of these two countervailing effects, it becomes interesting to understand the direction and magnitude of the interdependency between cumulative work and leisure readership.

In this paper, we examine the following three questions: 1) How do employees in an enterprise setting allocate their time to the different types of blogging activities when facing the trade-offs between work- and leisure-related content and between posting and reading content? 2) Is there any interdependence between work- and leisure-related readership accruing from blogging dynamics, and, if so, are these spillovers positive or negative in nature? 3) Would prohibiting leisure related blogging increase or decrease work related blogging in an enterprise? We formulate an employee's decision on when to write or read and what to write or read (in terms of work-related or leisure-related) as a dynamic game in the tradition of structural dynamic

---

[2]In addition, the CEO of the Fortune 1000 firm that provided us the data has stated on record that "[e]mployees' blogging activities are strongly correlated with their productivity."

games such as Erickson and Pakes (1995).

There are two main components of the employee's utility function, which forms the core of this paper. The first component is the cumulative readership of an employee's blog. In each period, an employee receives utility from her cumulative readership, separately for work- and leisure-related posts. Individuals also derive utility from the "knowledge state" they are in, which can be improved by reading blog posts (Singh et al., 2010; Yardi et al., 2009; Huh et al., 2007). This knowledge state forms the second component. Similar to the readership-based utility, we allow the knowledge-based utility derived from work- and leisure-related posts to be different.[3]

We apply the model to a unique dataset that comprises the complete details of blog posting and reading behavior for a large dataset of 2389 employees over a 15-month period at a Fortune 1000 IT services and consulting firm. Our results show that employees derive approximately 27% to 71% greater utility from cumulative work readership than from the same amount of cumulative leisure readership depending on the level of readerships. While readership of leisure posts provides less direct utility compared to work posts, employees still create leisure-related posts, as there is a significant spillover effect on the readership of work posts from the creation of leisure posts. We find that if there were no readership spillover effect from work to leisure, leisure posting would reduce by approximately 33.46%. Similarly, if there were no readership spillover effect from leisure to work, work posting would reduce by approximately 90.59%.

On average, we find that reading and writing work-related posts is more costly than reading and writing leisure-related posts. We also find that the cost of reading work and leisure posts together is lower than the sum of reading the two posts separately. This is potentially because of reduction of search costs. From the data, we note that in high fraction of instances readers read work and leisure posts from the same blogger. We also find a positive spillover effect between reading and posting blogs. By reading blogs individuals gain knowledge and bloggers with higher knowledge receive greater readership for their posts.

Competition among employees to attract readership for their posts leads to interesting tension for an individual as a blogger and as a reader. As a blogger, the individual would like peers to produce fewer posts so that she can gain higher readership. In contrast, as a reader she would like peers to produce more posts as she can gain more knowledge from them. We find that even a 10% decrease in competition could raise work posting by approximately 18% and work reading by approximately 13%.

Using the point estimates of the parameters, we implement a counterfactual experiment to analyze the effects of prohibiting leisure-related posting. Interestingly, this policy does not help increase work-related posting. Instead, it reduces the probability of work-related posting from 6.5% to 0.65% percent. This suggests

---

[3]Note that in this paper, an employee's knowledge state measures the knowledge he/she gains from blogging internally. The knowledge he/she achieves from external offline activities are not included since we do not observe that in our data.

that a policy of prohibiting leisure-related activities can hurt knowledge sharing in an enterprise setting. Intuitively, two countervailing forces exist under such a policy. On the one hand, a policy of disallowing leisure posting can provide employees with more time to generate work posting. On the other hand, this policy can also eliminate spillover effects in readership from leisure to work posting and, hence, reduce employees' incentives to create work-related posts. In our setting, the second force dominates the first, leading to an overall reduction in work-related posting.

Overall, our study aims to make the following contributions. First, in online settings, users need to allocate resources between content-generation and content-usage activities since they can take on the dual role of creators as well as consumers (Trusov et al. 2010; Ghose and Han 2011). However, there is little research that quantifies the interdependencies between how content on a social-media platform is created and consumed. The primary reason for such a gap is that, while data on content creation are easily available, data on content consumption are typically not available to researchers. However, the consumption information can only be retrieved from proprietary access logs. A small but emerging stream of work has begun to look at both content-creation and -consumption data, and our paper makes a significant contribution to this literature by providing insights into the tradeoffs between work- and leisure-related blogging and between content creation (blog writing) and content consumption (blog reading).. *Second,* to our knowledge, ours is the first study that provides a dynamic structural framework to analyze employee blogging activities.While prior studies have investigated why individuals create content on blogs, 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 to model employee behavior accordingly. Further, a key benefit of structural models is that they do not suffer from Lucas critique and they can be used to analyze the impact of policy interventions (Lucas 1976). As a result, our framework can be used for analyzing potential impact of policy interventions. *Third,* we demonstrate the existence of a positive spillover between cumulative work and cumulative leisure readership. Our model thus generates important managerial implications for firms that may be contemplating prohibiting leisure-related blogging within the enterprise. By demonstrating that there are positive spillovers from leisure-related blogging to work-related blogging and vice-versa, our results suggest that a policy of abolishing leisure-related content creation can have unintended adverse consequences on work-related content creation. *Finally*, a methodological contribution of our study is providing a way to account for unobserved states in Weintraub et al. (2008) framework through the use of a hidden Markov model (HMM) framework. We combine Weintraub et al. (2008) with the Arcidiacono and Miller (2011) framework to benefit from the computational simplicity of one approach and the ability to account for unobserved state variables of the other approach.

The rest of paper is organized as follows. In Section 2, we discuss the literature relevant to this paper.

Section 3 presents our model of employee blogging dynamics. Section 4 describes our data, followed by estimation strategies. We report estimation results in Section 5, followed by policy experiment and post hic analysis in Section 6. Section 7 summarizes and concludes our study.

2. Literature Review

Our research is related to multiple streams of literatures. The first stream of relevant literature relates to the impact of blogs. Several studies have focused on how different aspects of user-generated blogs affect product sales and market structure. Dewan and Ramaprasad (2012) show that the intensity of music sampling is positively associated with the popularity of a blog among previous consumers and that this association is stronger in the tail compared to the body of music sales distribution. Sun and Zhu (2012) analyze blog posts on a Chinese portal site after the launch of an ad-revenue-sharing program and find that popular content increases on blog posts after the introduction of the program. Hofstetter et al. (2009) study the causal effect of blog content on users' ability to form social ties and find that social ties affect content creation and vice-versa.

There are also quite a few papers investigating the effect of enterprise social media on firms' performance. For example, Huh et al. (2007) reveal that blogs facilitate 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. Studying a large internal corporate blogging community using log files and interviews, Yardi et al. (2009) find that employees expect to receive attention when they contribute to blogs, but these expectations often are not met. Singh et al. (2010a) study blog reading dynamics of employees within a large firm and find that most of the employees' time is devoted to reading and writing leisure-related posts. Aggarwal et al. (2010) find that negative blog posts act as a catalyst in increasing the overall readership of blogs, and the effect from this increased readership of positive posts is generally enough to offset the adverse effect of a few negative posts on firm value. Wattal et al. (2010) study the role of network externalities on the use of blogs in an organization and show that usage of blogs within an individual's network is associated with an increase in one's own usage and that this network effect is stronger for younger generations.

To our knowledge, no prior work has investigated the incentives for individuals to participate in blogging activities and examined the outcomes of firms' internal social-media policies. In online settings, users need to allocate resources between content-generation and content-usage activities since they can take on the dual role of creators as well as consumers (Trusov et al. 2010; Ghose and Han 2011). However, there is little research that quantifies the interdependencies between how content on a social-media platform is created and consumed. The primary reason for such a gap is that, while data on content creation are easily available, data on content consumption are typically not available to researchers. However, the consumption

information can be retrieved from proprietary access logs. A small but emerging stream of work has begun to look at both content-creation and -consumption data (Ghose and Han 2010, 2011; Ahn et al. 2011; Albuquerque et al. 2012). Ghose and Han (2010) find evidence of dynamic learning in multimedia content created by the mobile internet. In a dataset encompassing users' content creation and consumption behavior on the mobile internet, Ghose and Han (2011) find that there exists a negative interdependence between the content generation and usage behavior for a given user. Using an "approximation aggregation" rational-expectations equilibrium framework, Ahn et al. (2011) find that enhancing content durability and reducing content consumption cost appear to be the most effective strategies for increasing site visitation. Albuquerque et al. (2012) use data from a print-on-demand service of user-created magazines and find that content price and content creators' marketing actions each has a strong effect on purchases.

With regard to the methodology, our paper follows the literature on dynamic structural models. We use a dynamic game model to capture the key features of individuals' blogging behavior. The dynamic game model, in which multiple agents make decisions and the utility each agent gets depends on others decisions, is one specific type of dynamic structural model. It has been widely adopted and applied in I.O. and marketing (see Dube et al. 2009 for a detailed review). Pakes and McGuire (1994) and Ericson and Pakes (1995) first introduced Markov Perfect Equilibrium (MPE) as the equilibrium concept of dynamic games, and this equilibrium concept has been widely adopted since. However, the estimation of the dynamic game model is computationally burdensome. In fact, it MPE is computationally intractable for dynamic games with very large number of agents similar to ours. Recently, Weintraub et al. (2008) proposed the concept of oblivious equilibrium (OE) which approximates MPE under fairly general restriction. Another, challenge in dynamic game estimation dealt with accounting for unobserved heterogeneity. Recently, Arcidiacono and Miller (2011) proposed a strategy employing mixture model framework to account for unobserved heterogeneity in dynamic game models.

The most closely related studies to our work are Kumar and Sun (2009), who use a dynamic game to examine why users contribute to connected goods in social-networking sites, and Lu et al. (2010), who study how the social structure of individuals on a social-media platform affects their willingness to share knowledge with peers. However, none of these papers examines the implications of a firm's adoption of enterprise social media for internal employee behavior; nor do they examine employees' incentives to create and consume content internally. In our study, we use a dynamic game model to answer these questions, and this is among the first papers to adopt a structural-modeling approach to these questions.

3. Model

3.1. Per-Period Utility

Employees $i=1,\ldots, I$ decide about blogging decisions on a periodic basis for time periods $t=1,\ldots, T$. We define a period as one week. In enterprise blogging there are two types of blog posts that employees can generate represented by $j=\{w, l\}$ where w represents work-related posts and l represents leisure-related posts. In each period, an employee decides whether to read (or post a blog of type j). We use p to indicate 'post' and r to indicate 'read'. In other words, the action that an employee takes in each period is composed of four discrete elements, i.e. $d_{itwp}, d_{itlp}, d_{itwr}$ and d_{itlr}, where d_{itjp} is an indicator variable, which equals 1 if employee i posts one type j post at time t; and d_{itjr} is a count variable, which takes values 0, 1, 2 and 3. We allow individuals to post up to one blog in each period, but read up to 3 blogs in each period. If an individual posts more than one blog or read more than 3 blogs, the number will be truncated to the maximum number.[4] Hence, in total, there are 64 possible combinations of choices an employee can make. For notational convenience, we convert the four-dimensional action space to a one-dimensional action space A_i, which is defined as $A_i = \{1, 2, \ldots 64\}$, a finite set of 64 elements. In every period, every employee chooses an action $a_{it} \in A_i$. In addition, $a_t = (a_{1t}, \ldots, a_{It})$ denotes the set of actions that all employees choose at time t.

Note that each value of a_{it} is associated with only one combination of the four activities. For instance, $a_{it} = 1$ corresponds to the situation where $d_{itwp} = 0, d_{itlp} = 0, d_{itwr} = 0$ and $d_{itlr} = 0$; $a_{it} = 2$ indicates the situation where $d_{itwp} = 0, d_{itlp} = 0, d_{itwr} = 0$ and $d_{itlr} = 1$ etc. In other words, knowing a_{it} is equivalent to knowing $(d_{itwp}, d_{itlp}, d_{itwr}, d_{itlr})$. We model an employee's per period utility function at time t to comprise of utility from cumulative readership (denoted by R), utility from knowledge gained through blogging activities (denoted by K), an unobserved private shock, and everything else in the form of an outside good. An employee's utility at time t, U_{it}, is given by:

$$U_{it} = \omega_{it}(R_{itw}, R_{itl}, \theta_1, \theta_2) + \tau_{it}(K_{itw}, K_{itl}, \theta_3, \theta_4) + O_{it} + \gamma_{it}(a_{it}). \tag{1}$$

Here ω_{it} denotes cumulative readership-based utility and τ_{it} denotes knowledge-based utility from blogging. R_{itj} is the discounted cumulated readership employee i receives from type j posts up until period t and θ_1 and θ_2 are corresponding parameters that capture the effect of readership on utility. K_{itj} is the blog-related knowledge level of type j for an employee i at the end of time period t and θ_3 and θ_4 are corresponding parameters that capture the effect of knowledge on utility. O_{it} indicates the consumption of outside goods, with the utility from a per unit consumption of the outside good being normalized to one. The

---

[4]This assumption is based on the descriptive statistics of the data. It is motivated by the fact that fewer than 0.1% of users post more than 1 blog or read more than 3 posts of type j. Further, it is also necessary for computational tractability.

outside good option denotes different kinds of activities that an employee can engage in besides blogging. $\gamma_{it}(a_{it})$ is the action specific random shock associated with the utility that may affect an employee's decisions. Before choosing his/her actions, employee i receives a vector of choice-specific shocks, $\gamma_{it} = (\gamma_{it}(0), \gamma_{it}(1), ..., \gamma_{it}(63))$. To achieve identification, we assume that each element in γ_{it} has type one extreme value distribution and is *i.i.d.* across individuals and actions. When employee i chooses an action a_{it}, the choice-specific shock associated with this particular action, i.e. $\gamma_{it}(a_{it})$, is realized and goes into individual's current period utility. A summary of all notations and variables is given in Table 1.

3.1.1. Cumulative Readership

Prior work has shown that individuals who are identified as possessing expertise are often afforded power and status within the organization (French and Raven 1959). Hence, expertise sharing can produce significant personal benefit in terms of increase in recognition within the organization (Constant et al. 1994, Thomas-Hunt et al. 2003, Lu et al. 2010). These benefits are applicable in blog settings as well, since intuitively, when employees blog they are sharing their expertise with others. The readership of a blog indicates the extent to which the blogger shares her expertise and readers consider her an expert (Nardi et al 2004). Hence, the utility a blogger derives from posting would be proportional to the cumulative readership of her posts. In the utility function, the readership-based utility is incorporated as ω_{it} where

$$\omega_{it}(R_{itw}, R_{itl}, \theta_1, \theta_2) = \theta_1 log (R_{itw}) + \theta_2 log (R_{itl}) \tag{2}$$

We argue that employees derive readership-based utility from their cumulative readership (accumulated over time), and not just the contemporaneous readership (readership they receive in the current period). This is because employees' readership in previous periods can carry over to the next period with a certain discount rate. Even if one does not write any posts in the current period, he/she can still enjoy the benefits from the readership he/she has built in the past. We apply a log transformation here to adjust the over dispersion in the distribution of the cumulative readership. The log transformation also results in a concave relationship between readership-based utility and cumulative readership. This makes sense because one additional unit of readership does not provide as much additional utility to those who have a very high level of cumulative readership compared to those who have a low level of cumulative readership. We allow work and leisure readerships to enter separately into the model as they may lead to different kinds of incentives for blogging.

Table 1: Summary of Notations

$log(R_{itw})$	Natural log of cumulative readership (measured by depreciated past readership and current period readership) of work-related posts for employee i in period t.
$log(R_{itl})$	Natural log of cumulative readership (measured through depreciated past readership and current period readership) of leisure-related posts for employee i in period t.

K_{itw}	Work-related knowledge state from blogging (measured in levels). The number of levels specified after HMM estimation.
K_{itl}	Leisure-related knowledge state from blogging (measured in levels). The number of levels specified after HMM estimation)
d_{itwp}	Binary variable with 1 denoting employee i posts work-related post in period t and 0 otherwise.
d_{itlp}	Binary variable with 1 denoting employee i posts leisure-related post in period t and 0 otherwise.
d_{itwr}	Number of work-related posts employee i read in period t.
d_{itlr}	Number of leisure-related posts employee i read in period t.
	Other notations
i, j, t	Indices of employee, post types and periods (week)
a_{it}, s_{it}	Employee i's action and states in period t
a_t, s_t	Vectors of actions and states of all employees in period t
r_{itj}	New readership employee i receives in period t from type j posting
β, δ	Discount factors in life time utility function and depreciation factor in accumulation of the readership
$orgstatus_i$	Individual i's organizational status within the firm. It equals 1 if individual i's organizational status is high, and 0 if individual i's organizational status is low

3.1.2. Knowledge from Blogging

Blogs facilitate access to tacit knowledge and resources vetted by experts. The primary reason for corporations to allow their employees to participate in blogging activities during their work hours is that 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). Employees can acquire knowledge by reading other's posts. When employees read others' work-related posts they can become more productive, more informed about new ideas, and more aware of their colleagues' expertise, all of which opens up new opportunities for collaborations and for expressing their own opinions in future posts (Singh et al 2010). Employees can also gain knowledge by posting their own blogs. When employees create their own posts they have to synthesize different information, analyze it and then present it in a way that readers would understand. Similarly, reading or writing leisure posts can increase an employee's non work related knowledge. Further, individuals have an inherent need for leisure, which leisure reading and posting can provide. As in the case of readership, the two types of posts affect the type and level of incentives differently. Hence, we incorporate them separately. In the utility function, the blog knowledge-based utility is captured by τ_{it} where

$$\tau_{it}(K_{itw}, K_{itl}, \theta_3, \theta_4) = \theta_3 K_{itw} + \theta_4 K_{itl} \qquad (3)$$

3.1.3. Budget Constraint

Every week an employee has a limited amount of time, y_{it}, to engage in different activities. Note that y_{it} is

24×7 hours per week. Let t_{jrm} be the cost of identifying and reading m type j post and t_{jp} be the cost of developing and writing a type j post. Further, let t_{rr} (t_{pp}) represent any cost benefit by reading both leisure and work posts in same period; t_{wrp} (t_{lrp}) represent any cost spillover between work (leisure) writing and posting. Then we have the following budget constraint:

$$y_{it} = \sum_{j} t_{jp} d_{itjp} + \sum_{j} \sum_{m=1}^{3} t_{jrm}(d_{itjr} = m) - t_{rr} \log(d_{itwr} * d_{itlr} + 1) - t_{wrp} \log(d_{itwr} * d_{itwp} + 1)$$

$$- t_{lwp} \log(d_{itlr} * d_{itlp} + 1) - t_{pp} \log(d_{itwp} * d_{itlp} + 1) + t_o O_{it}.$$

$$(4)$$

Here, O_{it} is the outside good consumption and t_o is the associated coefficients that capture 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 between blogging and non-blogging activities. Let us define $\theta_5 = t_{wp}/t_0$; $\theta_6 = t_{lp}/t_0$; $\theta_7 = t_{wrm}/t_0$; $\theta_8 = t_{lrm}/t_0$; $\theta_9 = t_{rr}/t_0$; $\theta_{10} = t_{wrp}/t_0$; $\theta_{11} = t_{lrp}/t_0$; $\theta_{12} = t_{pp}/t_0$, which can be interpreted as the relative cost of participating in the four kinds of blogging activities compared to participating in non-blogging activities. Note that θ_7 and θ_8 are both 3x1 vectors. Solving for O_{it}, we get

$$O_{it} = \frac{y_{it}}{t_o} - \theta_5 d_{itwp} - \theta_6 d_{itlp} - \sum_{m=1}^{3} \theta_{7(m)}(d_{itwr} = m) - \sum_{m=1}^{3} \theta_{8(m)}(d_{itlr} = m)$$

$$+ \theta_9 \log(d_{itwr} * d_{itlr} + 1) + \theta_{10} \log(d_{itwr} * d_{itwp} + 1) + \theta_{11} \log(d_{itlr} * d_{itlp} + 1)$$

$$+ \theta_{12}(d_{itwp} * d_{itlp} + 1)$$

$$(5)$$

Combining equations (5) and (1) gives:

$$U_{it} = \theta_1 \log(R_{itw}) + \theta_2 \log(R_{itl}) + \theta_3 K_{itw} + \theta_4 K_{itl} + \frac{y_{it}}{t_o} - \theta_5 d_{itwp} - \theta_6 d_{itlp}$$

$$- \sum_{m=1}^{3} \theta_{7(m)}(d_{itwr} = m) - \sum_{m=1}^{3} \theta_{8(m)}(d_{itlr} = m) + \theta_9 \log(d_{itwr} * d_{itlr} + 1)$$

$$+ \theta_{10} \log(d_{itwr} * d_{itwp} + 1) + \theta_{11} \log(d_{itlr} * d_{itlp} + 1) + \theta_{12}(d_{itwp} * d_{itlp} + 1)$$

$$(6)$$

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

$$U_{it} = \theta_1 \log(R_{itw}) + \theta_2 \log(R_{itl}) + \theta_3 K_{itw} + \theta_4 K_{itl} - \theta_5 d_{itwp} - \theta_6 d_{itlp} - \sum_{m=1}^{3} \theta_{7(m)}(d_{itwr} = m)$$

$$- \sum_{m=1}^{3} \theta_{8(m)}(d_{itlr} = m) + \theta_9 \log(d_{itwr} * d_{itlr} + 1) + \theta_{10} \log(d_{itwr} * d_{itwp} + 1)$$

$$+ \theta_{11} \log(d_{itlr} * d_{itlp} + 1) + \theta_{12}(d_{itwp} * d_{itlp} + 1)$$

<div align="right">(7)</div>

Note here, y_{it}, the total time available is exogenous, which is 24×7 hours per week. However, the blogging budget is endogenous. One of the many (64) choices available to employees includes 0 work-related posting, 0 leisure-related posting, 0 work-related reading and 0 leisure-related reading. If an individual chooses this option, he/she spends the entire time budget on the outside good, and none on blogging activities. If an individual chooses to post both work-related and leisure-related blogs and read multiple work-related posts and multiple leisure-related posts, he/she spends a lot of time on blogging and less time on the outside good. In other words, by including the possibility of 0 posting and 0 reading in the action space, we are able to capture the trade-off individuals make between blogging and the spending time on the outside good.

3.2. Dynamic Game

3.2.1. State Variables

We define $s_{it} = (R_{itwp}, R_{itlp}, K_{itw}, K_{itl}, orgstatus_i)$ as the set of the state variables for an employee i at period t. We further define $s_{-it} = (R_{-itw}, R_{-itl}, orgstatus_{-i})$ as the set of observed state variables of i's peers. Individuals do not observe peers' knowledge states. This is a reasonable assumption because it is unlikely that individuals would know of the extent of knowledge gain from blog reading by others. In Section 5.5, we discuss how the data also supports this assumption. Then the strategy profile for i depends on $s_t = (s_{it}, s_{-it})$.

3.2.1.1. Cumulative Readership Evolution

The first four elements in the utility function (Equation 7) are all state variables that evolve according to the actions employees take in each period. The readership states evolve as follows:

$$R_{itj} = \delta R_{it-1j} + d_{itjp} r_{itj}.$$

<div align="right">(8)</div>

Here, δ is a depreciation factor, which is set at 0.95.[5] This depreciation factor accounts for the fact that the contribution of past cumulative readership to the current cumulative readership declines as time passes by. Individuals will have to continue posting to maintain a high cumulative readership. r_{itj} is the readership that blogger i receives for the type j post she wrote in period t and is determined as follows:

$$
\begin{aligned}
r_{itj} \\
= f \left(\begin{array}{c}
R_{itw}, R_{itl}, K_{itw}, K_{itl}, orgstatus_i, \sum_{x:x\neq i} d_{xtwp}, \dfrac{\sum_{x:x\neq i} d_{xtwp} R_{xtw}}{\sum_{x:x\neq i} d_{xtwp}}, \\[2mm]
\dfrac{\sum_{x:x\neq i} d_{xtwp} R_{xtl}}{\sum_{x:x\neq i} d_{xtwp}}, \dfrac{\sum_{x:x\neq i} d_{xtwp} orgstatus_x}{\sum_{x:x\neq i} d_{xtwp}}, \sum_{x:x\neq i} d_{xtlp}, \dfrac{\sum_{x:x\neq i} d_{xtlp} R_{xtw}}{\sum_{x:x\neq i} d_{xtlp}}, \dfrac{\sum_{x:x\neq i} d_{xtlp} R_{xtl}}{\sum_{x:x\neq i} d_{xtlp}}, \\[2mm]
\dfrac{\sum_{x:x\neq i} d_{xtlp} orgstatus_x}{\sum_{x:x\neq i} d_{xtlp}}, \sum_{x:x\neq i} d_{xtwr}, \sum_{x:x\neq i} d_{xtlr}, \varepsilon_{itj}
\end{array} \right)
\end{aligned}
\tag{9}
$$

The readership of a new type j post is affected by a blogger's own cumulative work and leisure readerships, knowledge levels and her organizational status. It is also affected by the number of work and leisure posts written and posted by her peers in the period t, and the average work and leisure readerships and organizational status of peers' who posted work and leisure posts in period t. These variables that capture peer posting activity account for the competition effect. The amount of reading represents the size of readership pie and the amount of posting represents the number of individuals among whom the pie will be distributed.

Further, to account for unobserved or latent factors that may affect r_{itj} we include a term ε_{itj}. An example of this latent factor is the timeliness of one's blog postings. A highly timely post may acquire a higher readership for its poster. Note that posting is a necessary but not sufficient condition for getting readership. As described earlier there are a number of variables that affect the readership of a post.

3.2.1.2. Evolution of Knowledge From Blogging

The third and fourth state variables are the blog-related knowledge levels of employee i, K_{itw} and K_{itl}. The knowledge of an employee is unobservable to us and her peers. We model that work knowledge has $1,\ldots, kw$ ordered discrete levels and the leisure knowledge has $1,\ldots, kl$ ordered discrete levels. In period t, an individual probabilistically belongs to a knowledge state based on her knowledge stock, 1 being the lowest knowledge state and $kj; j \in \{w, l\}$ being the highest. She can transition from one state to another through participation in learning activities, some of which are observed, over a given period. This allows for the transition probability from a given state to another to be different across individuals as well as across time periods for the same

---

[5]The qualitative nature of the results is robust to several other values of the depreciation factor as shown in the section on robustness checks.

individual. In our setting, the two learning activities we observe are the amount of reading and posting activities of an employee in a period. The probability of individual i to transition from knowledge state K_{it} at time t to K_{it+1} at time $t+1$ is given as

$$P\big(K_{it+1j} = k_{t+1}|K_{itj} = k_t, \, s_{it}, \, s_{-it}\big)$$
$$= f\left(K_{itj}, d_{itjp}, d_{itjr}\left(\sum_{x:x \neq i} d_{xtjp}\right), d_{itjr}\left(\frac{\sum_{x:x \neq i} d_{xtjp} \, R_{xtj}}{\sum_{x:x \neq i} d_{xtjp}}\right), \frac{\sum_{x:x \neq i} d_{xtwp} \, orgstatus_x}{\sum_{x:x \neq i} d_{xtwp}}, \kappa_{itj}\right)$$

(10)

Thus, we model the type j knowledge state transitions to depend on both reading and posting activities of the same type. The state transition also depends on the number of type j posts written by peers and the average type j cumulative readership and organizational status of peers who wrote those posts in period t. The knowledge state of an individual would influence her ability to learn (Cohen and Levinthal 1990). Reading and posting are learning activities which should affect knowledge state transitions. The number of posts written by peers represents the richness of the content from which a reader could learn. Further the posts written by bloggers with higher cumulative readership and organizational status are likely to be of higher quality influencing the learning of the readers. Finally, knowledge state transition could also be affected by the hidden or latent factors unobserved to us which are captured by the term κ_{itj}.

3.2.2 Sequence of Events

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

1. Employees observe their states s_{it} and their peers' states s_{-it}.
2. Employees receive a set of choice-specific random shocks (γ_{it}).
3. Employees calculate their expected readership and knowledge states as a function of potential decisions they and their peers can make.
4. Employees execute their decisions and receive utility.
5. Employee states evolve to s_{it+1} because of their and their peers' decisions.

3.2.3. Long Term Utility Function

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

$$\max_{a_{it}} E\left(\sum_{t=1}^{\infty} \beta^t U_{it}(a_{it}, s_t, \gamma_{it})|s_t, \gamma_{it}\right)$$

(11)

Here, β 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 three components of the model that need to be emphasized. An employee in our model maximizes his/her lifetime utility, which makes the model dynamic. Since, the present actions and states of an employee affect his/her future utility by affecting the states, the employee is forward-looking. The employee balances his/her time between reading and writing, and work and leisure due to the budget constraint. Hence, his/her actions are interdependent. At the same time, the utility of an employee is a function of the decisions made by his/her peers (through readership and knowledge state transitions) making it a multi-agent dynamic game.

3.2.4. Equilibrium concept

Following Ericson and Pakes (1995) we focus on MPE as a solution concept. 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 $\sigma_i: S \times \Gamma_i \rightarrow A_i$, where S denote states of all individuals, Γ_i denotes employee i's private random shock, and A_i denotes the set of individual i's actions. A profile of Markov strategies is a vector, $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_I)$, where $\boldsymbol{\sigma}: S \times \Gamma_1 \times \ldots \times \Gamma_k \rightarrow A$. Here, k is the number of individuals and we drop the time index because the strategy profile is time invariant. If behavior is driven by a Markov strategy profile $\boldsymbol{\sigma}$, employee i's expected utility given state $\boldsymbol{s}$ can be written recursively as Bellman Equation:

$$V_i(\mathbf{s}; \boldsymbol{\sigma}) = E_\gamma \left[U_i(\boldsymbol{\sigma}(\mathbf{s}, \boldsymbol{\gamma}), \mathbf{s}, \gamma_i) + \beta \int V_i(\mathbf{s}'; \boldsymbol{\sigma}) dP(\mathbf{s}' | \boldsymbol{\sigma}(s, \boldsymbol{\gamma}), s) \Big| \mathbf{s} \right].$$

(12)

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

$$V_i(\mathbf{s}; \boldsymbol{\sigma}, \boldsymbol{\sigma}_{-i}) \geq V_i(\mathbf{s}; \sigma_i', \boldsymbol{\sigma}_{-i})$$

(13)

However, given the large number of agents and the huge state space, MPE is computationally intractable in our setting. Recently, Weintraub et al (2008) developed the concept of OE which approximates MPE under fairly general assumptions. Farias et al. (2012) find that if the market is not too concentrated and the number of agents is large, which is the case in our setting, OE approximates MPE very well.

OE is based on the idea that in a market with large number of agents, simultaneous changes in states of large number of actors would average out and the industry state would remain roughly constant over time (Weintraub et al 2008). As a result, an agent can make near optimal decisions by considering her own state and the long run average industry state. OE is also a more reasonable equilibrium concept in our setting as

individuals are unlikely to know states of each of their peers which they are required to do in MPE. Further, in our dataset we can see that the average peer states are relatively stable and are not trending. Hence, we use OE to approximate MPE in our setting.

4. Empirical Estimation

4.1. Data

Our data come from a large Fortune 1000 IT services, business-process outsourcing, and consulting firm. It is a U.S.-based firm with significant presence and operations in several other countries across multiple continents: Europe, Asia, and the Americas being the major areas. *Fortune* magazine named this firm one of the fastest-growing companies in 2009, with several billion dollars in annual revenue. 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 such as blogs for use within the enterprise.

These blogs are hosted on an internal platform and are not accessible by people outside the organization. On this platform, every employee is allowed to host his/her own blog, which is accessible to all of the firm's employees across the entire hierarchy. The identity of the blogger is revealed on the blog. A brief description of employees' most recent posts is displayed on their blog homepages in the order in which they were posted (the most recent post appears at the top). This listing is done irrespective of the category or topic of the post. That is, both work- and leisure-related posts appear on the homepage. The titles of these posts act as hyperlinks on which the users can click to get to the full content of the post. Notably, the firm offers no explicit reward structure.

Bloggers classify their posts into one of several categories (for example, software testing, movies, history, knowledge management, senior management, etc). To be able to measure the knowledge-sharing aspect, the firm tracks which employee reads which blog at what time; the readership count of each post is displayed along with the blog post and is observable to everyone. Since the blogs are only internally accessible, the firm does not impose any restrictions on the kind of posts that employees can write. 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 2 presents the sub-categories that constitute each topic.

Table 2: Blog-Post Classification

Topics	Leisure-Related	Work-Related

Sub-Categories	Fun; Movies-TV-Music; Sports; Puzzles; Chip-n-putt; Religion-Spiritual-Culture; History-Culture; Photography; Arts; Poetry-Stories; Books; Geographies	FLOSS; Technology; Testing; Domains; Corporate Functions; Knowledge Management; Project Management; Business Development; Senior Management; Practices-Programs-Accounts

Our data consist of the blog reading and writing activities of 2389 employees over a 15-month (64 weeks) period. These include employees who both post and read at least one post during the sample period. 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 periods in total. We treat the first 16 weeks as the hold-out period for calculating initial state variable values. We estimate and test the model using data from the remaining 48 weeks. [6] We use the data from week 17 to 56 for estimating the model parameters. We use data from week 57 to 64 for testing the model performance.

Table 3 contains high-level descriptive statistics of our data. These employees wrote 25,981 posts during the 64-week period, indicating that not every employee generated a blog posting in every period. Of these, 9,934 posts were work-related and 16,047 leisure-related. There were 26,784 readings of work posts and 37,232 readings of leisure posts by these employees during the 64-week period. We classify the organizational status of an employee into "high" and "low" based on his position in the organizational hierarchy. Because this firm is in the IT industry, anyone at or above the rank of a Director in the company is classified as a "high-status" individual. Of the 2389 employees in our sample, 210 employees are classified as "high-status," while the remaining 2,179 are classified as "low status."

The descriptive statistics for the key variables used to construct the model variables are presented in

Table **4**. On average, every week, 6.50 percent of employees post a work-related blog; 10.50 percent post a leisure-related blog; 17.52 percent read work-related blogs; and 24.35 percent read leisure-related blogs. 40.03 employees, on average, read a given work-related post, while 107.48 employees, on average, read a given leisure-related post.

Table 3: Overall Sample Statistics

Number of Employees	2389
Number of High-status Employees	210

---

[6] 45,287 employees participated in blogging over the 64-week period. However, most of these employees were not active users of blogging. Only 2389 employees created at least one post and read at least one post in the 64-week period. These 2389 employees constitute the full set of employees who ever wrote a post. However, these employees form only a subset of the total readership. In the paper, we model the reading and posting decisions of these 2389 employees. However, because all employees know how much readership each post received, the readership equation in the paper uses the true readership data (from the full sample of 45,287 potential readers). We have also conducted a robustness check, where we consider only readings by the 2389 employees in the estimation- of-readership equation. The qualitative nature of all our results remains the same.

Number of Low-status Employees	2179	
Period length	1 week	
Number of periods	64	
Total posts written	25981	
Work-related posts written	9934	
Leisure-related posts written	16047	
Work-related post reading[7]	26784	
Leisure-related post reading	37232	

Table 4: Descriptive Statistics of Key Variables[8]

Variable	Mean*	Standard Deviation	Minimum	Maximum
R_{itw}	16.233	85.18	0	1508.94
R_{itl}	86.114	291.13	0	1998.32
K_{itw}	0.119	0.33	0	1
K_{itl}	0.158	0.37	0	1
d_{itwp}	0.065	0.18	0	1
d_{itlp}	0.105	0.26	0	1
$d_{itwr} = 1$	0.080	0.16	0	1
$d_{itwr} = 2$	0.034	0.0.03	0	1
$d_{itwr} = 3$	0.008	0.03	0	1
$d_{itlr} = 1$	0.124	0.23	0	1
$d_{itlr} = 2$	0.037	0.04	0	1
$d_{itlr} = 3$	0.015	0.03	0	1
$orgstatus_i$	0.088		0	1

* We first calculate the average of each variable across individuals in each period and then calculate the mean across periods.

4.2. Estimation Strategy

Our model and data raises several challenges for estimation. First, the readership states are continuous. The continuous states make the estimation computationally very difficult. As a result, we discretize the log of cumulative readership states (from *0* to *7.7*) into *7* equal sized intervals. We chose *7* intervals as it allows us to have sufficient data in each discrete interval and have meaningful transitions from each state. This allows us to calculate the state transitions appropriately. Because, we discretized the readership states, we also discretize the corresponding parameters in the utility function. For each readership state we have *7* parameters one corresponding to each level of readership state. That is θ_1 and θ_2 are now vectors of size *7x1* each. The state size (*NS*) for an individual's own state is *7*7*2*kw*kl*. The size of the individual specific state transition matrix would be *NSxNS*. Second, knowledge states are unobserved. Hence, as mentioned earlier we use the

---

[7] The number refers to the number of times the 2389 individuals read posts during the entire 64 weeks. In Equation (9), new readership *r* includes readership from all employees in the company, not only the 2389 individuals.
[8] The descriptive statistics for the knowledge states provided here are based on the identification of the knowledge states which will be discussed later.

procedure proposed by Arcidiacono and Miller (2011) to handle models with unobserved state variables.

The estimation procedure works as follows.

Step 1. Estimate the transition probabilities for $\log(R_{itw})$, $\log(R_{itl})$, K_{itw}, and K_{itl} and the Conditional Choice Probabilities (*CCP*).

All the states have discrete ordered levels. As a result, we model the state transitions as an ordered logistic regressions. We allow the parameters of the regression to be current state specific. The *CCP* represents the probability of a choosing an action given the state values. Given the states we can estimate the *CCP* through a multinomial logistic regression of actions on state variables. The *CCP* and the state transitions are jointly estimated through a *HMM*.

The *HMM* works as follows: The state transition probabilities for $\log(R_{itw})$ and $\log(R_{itl})$ are a function of variables given in Equation 9. And the state transitions for K_{itw}, and K_{itl} are function of variables given in Equation 10. Let $D_i = \{a_{i1}, a_{i2}, \ldots, a_{iT}\}$ represent the sequence of choices for individual *i*; $S_i = \{s_{i1}, s_{i2}, \ldots, s_{iT}\}$ represent the state sequence over time for individual *i*, π_i represent the initial state distribution (*1xNS*) for individual *i*, λ represent the set of parameters that govern the state transition probabilities and the *CCP (NSx64)*. An element *jk* of *CCP* represents the probability of action *k* given state *j*. Then the probability of observing D_i given a state sequence and parameter set λ is given as

$$P(D_i|\lambda, S_i) = \prod_{t=1}^{T} P(a_{it}|\lambda, s_{it}).$$

(14)

Thus we obtain

$$P(D_i|\lambda, S_i) = P(a_{i1}|s_{i1}) * P(a_{i2}|s_{i2}) * \ldots * P(a_{iT}|s_{iT}),$$

(15)

where $P(a_{it}|s_{it})$ is the probability of observing action a_{it} given individual *i* is in state s_{it}. Note that $P(a_{it}|s_{it})$ is an element of *CCP* and given the parameters it is calculated using a multinomial logistic specification. The probability of state sequence S_i is given as

$$P(S_i|\lambda) = \pi_i * P(s_{i2}|s_{i1}) * \ldots * P(s_{it}|s_{it-1}) * \ldots * P(s_{iT}|s_{iT-1}).$$

(16)

Here $P(s_{it}|s_{it-1})$ is the probability of transitioning to state s_{it} at time *t* given state s_{it-1} at time *t-1*. Given the parameter set λ, and ordered logistic specification for state transitions, we can calculate the probabilities of transitions for $\log(R_{itw})$, $\log(R_{itl})$, K_{itw}, and K_{itl} individually conditional on the unobserved states. Then $P(s_{it}|s_{it-1})$ is obtained as follows:

$$P(s_{it}|s_{it-1})$$
$$= P(\log(R_{itw}), \log(R_{itl}), K_{itw}, K_{itl}, Orgstatus_i | \log(R_{it-1w}), \log(R_{it-1l}), K_{it-1w}, K_{it-1l}, Orgstatus_i, \lambda)$$
$$= P(\log(R_{itw})|s_{it-1}, \lambda) * P(\log(R_{itl})|s_{it-1}, \lambda) * P(K_{itw}|s_{it-1}, \lambda) * P(K_{itl}|s_{it-1}, \lambda).$$

$$(17)$$

Here, each of the probabilities is coming from separate ordered logistic specifications. For calculating $P(\log(R_{itj})|s_{it-1}, \lambda)$ we employ variables specified in Equation 9. For calculating $P(K_{itj}|s_{it-1}, \lambda)$ we employ variables specified in Equation 10. For simplicity let us represent the observed state by O, and unobserved/hidden state by . Further, let O_i and H_i represent the observed and unobserved state sequences for individual i, respectively.

Then we can write $P(s_{it}|s_{it-1})$ as

$$P(s_{it}|s_{it-1}) = P(O_{it}|H_{it-1}, O_{it-1}, \lambda) * P(H_{it}|H_{it-1}, O_{it-1}, \lambda) \tag{18}$$

And Equation 16 can be written as

$$P(H_i, O_i|\lambda) = \pi_i * \prod_{t=1}^{T} P(O_{it}|H_{it-1}, O_{it-1}, \lambda) * P(H_{it}|H_{it-1}, O_{it-1}, \lambda) \tag{19}$$

The probability that D_i and O_i occur simultaneously is given as

$$P(D_i, O_i|\lambda) = P(D_i|\lambda, O_i, H_i) * P(O_i|\lambda, H_i) * P(H_i|\lambda). \tag{20}$$

Hence, the probability of the observed outcome sequence D_i and O_i is obtained by summing Equation (20) over all possible values of unobserved state sequence H_i (Rabiner 1989), explicitly,

$$L(D_i, O_i) = P(D_i, O_i|\lambda) = \sum_{\forall S_i} \left(P(D_i|\lambda, O_i, H_i) * P(O_i|\lambda, H_i) * P(H_i|\lambda) \right). \tag{21}$$

We maximize Equation 21 to obtain parameter set λ. The initial state distribution is analytically derived by solving the equation $\pi_i = \pi_i \bar{Q}$, where $\bar{Q}$ is the state transition matrix calculated at the mean value of covariates for individual i. The number of levels for knowledge states is determined by comparing models with different number of levels for knowledge states.

Step 2. Calculate Utility Parameters

In this step we estimate the structural parameters: $\rho = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{11}, \theta_{12})$. In an OE, the *CCP* is function of individual's own state. Peers' influence on profit function of an individual is due to the fact that *CCP* along with state transition probabilities determine the long run industry state. The main computational advantage that OE provides is to convert the multi agent dynamic game problem to a single

agent dynamic optimization problem. And one can use any one of several methods proposed in literature to solve this single agent dynamic optimization problem. We follow Agurregabiria and Mira (2007)'s nested pseudo maximum likelihood procedure to solve the single agent dynamic optimization problem.

The estimation procedure works as follows.

Iterate through the following steps until $\rho^m \sim \rho^{m+1}$, where ρ^m is the value of structural parameters in the mth iteration.

1. Given a long run aggregate peer state distribution, parameter set λ, and *CCP*, calculate action specific state transition matrices (Q_a). The peer specific variables needed to calculate the state transition probability are calculated using the aggregate peer state distribution and *CCP*. This is possible because peers also follow the same *CCP*.

2. Using the *CCP* and the action specific state transition matrices calculate the unconditional state transition matrix (Q) as $Q = \sum_{a=1}^{64} CCP(.,a) * Q_a$. Here $CCP(.,a)$ is the column *a* of *CCP*.

3. Calculate the long term state specific value function V as

$$V = (I - \beta Q)^{-1} * \left[\sum_{\forall a} CCP(.,a)(u(a) + \gamma(a)) \right]$$

(22)

Here *I* is an identity *(NSxNS)* matrix, $u(a)$ is a *NSXnp* action specific expected profit matrix before the random shock is realized (*np* is the number of structural parameters to be estimated), $\gamma(a)$ is an *NSx1* action specific random shock vector, and represents an element by element multiplication. Note that $u(a)$ is current period utility without the random shock.

4. Calculate state and action specific long term value function V_a as

$$V_a = u(a) + \beta * Q_a * V.$$

(23)

5. Find structural parameter set ρ by maximizing *L(D_i)* from equation 10.

   -Given a value of ρ and V_a calculate *P(a|s)* as

$$P(a|s) = \frac{exp(V_a(s))}{\sum_{\forall a} exp(V_a(s))}.$$

(24)

   Here $V_a(s)$ is the long term value from taking action *a* for an individual in state *s*.

   -Use $P(a|s)$ calculated above, λ and π_i calculated in Step 1, to calculate *L(D_i)*.

   - Find ρ by maximizing *L(D_i)*.

6. Using ρ and V_a update the *CCP*.

7. Calculate the long run aggregate peer state distribution.

      -update Q using Q_a and *CCP*.

      -calculate *100* period state transition probability matrix Q^{100} (Weintraub et al 2008).

      -the long run aggregate peer state distribution is any row or column of the Q^{100}.

4.3. Identification and Normalization

There are a few identification issues that need to be addressed before the model can be consistently estimated. First, we assume that the private shocks are extreme value type 1 distributed as is common in the literature (e.g. Hotz and Miller 1993, Weintraub et al. 2008). Second, in the HMM we assume that the probability of knowledge increase is higher for an individual as a result of reading than otherwise. We allow this by assuming that reading has a non-negative impact on knowledge increase. Third, we cannot identify t_o and the cost-specific parameters together. Hence, we normalize t_o =1. Fourth, for identification we specify the utility parameter corresponding to the lowest level of readership state as zero. Fifth, we assume that the knowledge state of an employee is a private state and make an exclusion restriction. An employee's actions, while depending on his/her own knowledge state, are independent of his/her peers' knowledge states. Such exclusion restrictions are commonly imposed in dynamic structural models (Bajari et al 2009). In Section 4.5 we relax some of these assumptions to check for robustness.

      To ensure that only the data after the equilibrium is reached is used for estimation we use a part of our data (16 weeks) as the pre-estimation sample and estimate our model using the rest of the data (48 weeks). We also do a robustness check with a 30-week hold out period and find no change in the qualitative nature of the results.

5. Results

In Step one of the estimation, we identify the number of levels for the knowledge states. We compared several models with different number of knowledge state levels from 1 to 5. The results indicate that a model with two levels of knowledge for both work and leisure perform the best in terms of Akaike Information Criterion (*AIC*) as well as Bayesian Information Criterion (*BIC*). Hence, we conduct the remaining analysis with two levels for both work and leisure knowledge states. In step one, we also estimated the state transition probabilities. These transition probabilities reveal several insights which we discuss in the next subsection. Due to space limitations, we discuss only the work readership and knowledge transitions.

5.1. Readership State Transitions

5.1.1. Readership Spillover from Leisure to Work

In Table 5, we report the work cumulative readership state transitions from posting a work post at time t for a low organizational status individual that is in low knowledge state. For constructing this table we set the peer specific variables (from Equation 9) at their mean values. Leisure knowledge of the employee does not significantly influence the work readership transition.

Table 5: Expected Cumulative Work Readership State at Time t+1 due to Posting of One Work Post for Employee in Low Work Knowledge State *

		Cumulative Leisure Readership State at Time t						
		1	2	3	4	5	6	7
Cumulative Work Readership State at Time t	1	1.120	1.153	1.193	1.243	1.303	1.374	1.457
	2	2.116	2.155	2.196	2.246	2.307	2.379	2.464
	3	3.099	3.134	3.170	3.215	3.270	3.336	3.414
	4	4.084	4.115	4.146	4.186	4.234	4.294	4.365
	5	5.050	5.071	5.091	5.116	5.148	5.187	5.235
	6	6.017	6.029	6.037	6.048	6.061	6.078	6.099
	7	7.000	7.000	7.000	7.000	7.000	7.000	7.000

*low organizational status employee; peer specific variables are at mean level.

Table 5 shows that the expected work readership state is higher than an employee's current state if she posts a work related post. That is the new readership that a new work post will bring in outweighs the depreciation in existing readership. Interestingly, the expected work readership as a result of posting a work post is higher for employees which higher leisure readership than otherwise. For example, for an employee in work readership state of 4 the expected work readership state from posting a work post is 4.186 if her leisure readership state of 4 and 4.365 if her leisure readership state is 7. This indicates a positive spillover from leisure to work readership. As mentioned earlier, several posts of a blogger are displayed on the same homepage irrespective of whether they are work or leisure. As a result, when readers visit a blog they are exposed to both work and leisure posts. Hence, some readers who may visit the blog for reading a leisure post may also end up reading a work post. We also find similar spillover effect from work readership to leisure readership.

5.1.2. Spillover from Work Reading to Work Posting

In Table 6, we report the expected work readership state from posting a work post for a low organizational status employee who is in high work knowledge state. The peer specific variables used in constructing Table 6 are set at their mean values. Work reading primarily affects a reader's work knowledge state. To see the spillover effect from work reading to work posting, we can compare how work knowledge affects the work

readership state evolution. By comparing values in Table 5 to their corresponding values in Table 6, we can see that the values are higher in Table 6. For example, the expected work readership from posting a work post for an employee with work readership state of 4 and leisure readership state of 4 is 4.371 if she has high work knowledge and 4.186 if she has low work knowledge. This indicates there is a positive spillover from work reading to work posting. As discussed earlier, bloggers with greater knowledge may be able to write better quality posts which attract more readers. We also find similar spillover effect from leisure reading to leisure posting.

Table 6: Expected Cumulative Work Readership State at Time t+1 due to posting of one Work Post for Employee in High Work Knowledge State*

		Cumulative Leisure Readership State at Time t						
		1	*2*	*3*	*4*	*5*	*6*	*7*
Cumulative Work Readership State at Time t	*1*	1.247	1.308	1.380	1.464	1.561	1.669	1.787
	2	2.251	2.312	2.385	2.471	2.568	2.677	2.797
	3	3.219	3.275	3.342	3.421	3.513	3.618	3.735
	4	4.189	4.238	4.299	4.371	4.456	4.555	4.667
	5	5.119	5.151	5.191	5.239	5.298	5.368	5.450
	6	6.049	6.062	6.079	6.101	6.127	6.158	6.196
	7	7.000	7.000	7.000	7.000	7.000	7.000	7.000

*(low organizational status employee; peer specific variables are at mean level).

5.1.3. Effect of Competition on Work Readership Evolution

The readership state transition regressions show a negative and significant effect of peer posting related variables on an individual's transitions to higher readership states. To test how competition for readership affects the evolution of work readership states we calculate readership state transition probabilities at a higher level of competition. To conduct this analysis we increase the peer specific variables (other than number of posts read by peers, and the variables corresponding to peers' leisure posting) by 10% and recalculate the state transition probabilities and report the results in Table 7. Table 7 is constructed for a low organizational status employee in high knowledge state. In constructing Table 7 we assume peers wrote 10% more work posts than their average, average work and leisure readership state and organizational status of these peers is 10% higher than on average. This means more peers are writing work posts and these peers have higher work and leisure readership state and organizational status than the mean levels of peers who wrote posts in a period. All other peer specific variables that affect work readership evolution are at their mean levels.

On comparison of Table 7 with Table 6 we see that there is competition for readership. As competition increases the expected new readership that an employee's new work post can get decreases. For

example, the expected work readership state from posting a work post decreases from 4.371 to 4.163 when the competition increase by 10% from its mean level for a low organizational status employee in high work knowledge state whose work and leisure readership states are both 4. We find similar effect of competition on leisure readership state evolution.

Table 7: Expected Cumulative Work Readership State at Time t+1 due to posting of one Work Post when Competition is 10% more Intense*

		Cumulative Leisure Readership State at Time t						
		1	*2*	*3*	*4*	*5*	*6*	*7*
Cumulative Work Readership State at Time t	*1*	1.105	1.134	1.170	1.215	1.269	1.334	1.411
	2	2.064	2.136	2.173	2.218	2.273	2.339	2.416
	3	3.050	3.118	3.150	3.190	3.239	3.299	3.370
	4	4.037	4.101	4.128	4.163	4.207	4.260	4.325
	5	5.008	5.062	5.080	5.102	5.130	5.165	5.208
	6	5.980	6.025	6.033	6.042	6.054	6.068	6.087
	7	7.000	7.000	7.000	7.000	7.000	7.000	7.000

*low organizational status employee in high work knowledge state.

5.1. Knowledge State Transitions

To see how work reading affects work knowledge state transitions we report 4 transition matrixes in Table 8. For constructing the transition matrixes in Table 8 we assume the peer specific variables (from equation 10) are at their mean values and the employee has not written a work post in period *t*. Since the organizational status, and work and leisure readerships of the employee do not affect her work knowledge transitions we ignore them while constructing these transition matrixes. Table 8a represents a work knowledge transition matrix for an individual who has not read any work post in period *t*. Tables 8b, 8c, 8d represent the work knowledge transition matrixes for cases where an individual has read *1*, *2*, and *3* work posts in period *t*, respectively.

From Table 8a, we can see that the states are very sticky. It is hard to move from one state to another. The probability that an individual in low state may move to a high state is only 2.93% if he does not read or write a work post. However, reading work posts increases work knowledge of an employee. We can see from Table 8 that the reading 1, 2, and 3 work posts increases the probability of an employee in low work knowledge state to high work knowledge state from 2.93% to 4.83%, 6.12%, and 7.10% respectively. Reading work posts also helps an employee in high work knowledge state to stay in high work knowledge state.

Table 8: Work knowledge Transition Probability from Reading*

[Low Work knowledge (LWK); High Work Knowledge (HWK)]

	(a)	(b)	(c)	(d)

		Read 0 work post		Read 1 work post		Read 2 work post		Read 3 work post	
		State in period t+1		State in period t+1		State in period t+1		State in period t+1	
		LWK	HWK	LWK	HWK	LWK	HWK	LWK	HWK
State at period t	LWK	97.07%	2.93%	95.17%	4.83%	93.88%	6.12%	92.90%	7.10%
	HWK	15.45%	84.55%	9.80%	90.20%	7.80%	92.20%	6.73%	93.27%

*Peer related variables at mean value and zero work post written by individual.

5.2.1. Spillover from Posting to Reading

An individual can also learn while developing her blog post. As a result there could be spillover from work posting to work reading. Further readers may also benefit if there are more posts written by their peers. If there are more posts available to read a reader would have greater variety of content to choose from. Also, work posts by employees who have higher work readership or organizational status may represent higher quality content. Readers could benefit from access to such high quality content.

In Table 9a we present the state transition matrixes for an employee who posted one work post in period t and read zero work posts in period t. Comparing Table 9a with 8a, we can see that individual is able to gain work knowledge by writing a work post also. In fact, writing a work post provides greater increase in work knowledge compared to reading a work post. Specifically, by writing a work post an employee's probability of moving from low work knowledge state to high work knowledge state increases from 2.93% to 5.90%.

In Table 9b,9c, and 9d we present the work knowledge transition matrixes where an individual has read *1*, *2*, and *3* work posts respectively, and posted zero work post, and peers have written 10% more posts than on average, the average readership and organizational status of peers who wrote work posts is 10% higher than on average. Here, we can see that as peers write more posts and more high quality posts are available readers learn more. For example, the probability to transition from a low work knowledge state to a high work knowledge state increases from 4.83% to 6.17% when peers have written 10% more posts and the posts are written by individuals with 10% higher work readership states and organizational statuses than on average. This finding highlights the counter veiling effects of peer blogging. As we showed earlier, as a blogger an employee competes with her peers to attract readership to her posts. The value that a blogger may derive from her posts decreases when peers write more posts or when peers with higher readership or organizational status write posts. Hence, as a blogger an employee would prefer peers to write fewer posts. However, as a reader the same blogger benefits when her peers write more posts.

Table 9: Work Knowledge Transition Probabilities from Posting and Greater and Better Content

[Low Work knowledge (LWK); High Work Knowledge (HWK)]
(3a. zero work posts read by individual, 3b. zero work posts written by individual and peer related variables
are 10% higher than their mean values)

		(a)		(b)		(c)		(d)	
		Post 1 work post		Read 1 work post		Read 2 work post		Read 3 work post	
		State at period t+1		State at period t+1		State at period t+1		State at period t+1	
		LWK	HWK	LWK	HWK	LWK	HWK	LWK	HWK
State at period t	LWK	94.10%	5.90%	93.83%	6.17%	91.71%	8.29%	89.27%	10.73%
	HWK	8.13%	91.87%	7.65%	92.35%	5.46%	94.54%	4.81%	95.19%

5.3. Utility Function Parameters Results

The results for the second stage are presented in 10. For estimation purposes, the discount factor β is set to 0.95.[9] The results show that employees gain positive utility from work-related cumulative readership and leisure-related cumulative readership. Employees also gain positive utility from work and leisure knowledge.

Table 10: Estimated Utility Function Parameters

	Work Related Parameters	Leisure Related Parameters
Readership Related Parameters		
Cum Readership level =1	Fixed at zero	Fixed at zero
Cum Readership level =2	0.505 ***	0.331 ***
Cum Readership level =3	1.231 ***	0.890 ***
Cum Readership level =4	2.062 ***	1.621 ***
Cum Readership level =5	3.002 ***	2.153 ***
Cum Readership level =6	5.094 ***	3.294 ***
Cum Readership level =7	9.157 ***	5.357 ***
Knowledge Related Parameters		
Knowledge Level = Low	Fixed at zero	Fixed at zero
Knowledge Level = High	1.351 ***	0.116 **
Cost Related Parameters		
Writing One Post	5.699 ***	4.013 ***
Reading One Post	2.431 ***	1.703 ***
Reading Two Post	2.709 ***	2.260 ***
Reading Two Post	3.792 ***	2.721 ***
Cost Interaction Parameters		
Reading X Posting	-0.113 *	-0.035
Work Posting X Leisure Posting	-0.019	
Work Reading X Leisure Reading	-0.249 ***	

Comparing the parameters corresponding to the readership state variables, we can see that same level of work readership provides greater utility compared to leisure readership. For example, an individual in work readership level *4* gets a one period utility of *2.062* from her being in work readership level *4*. In comparison,

---

[9]The qualitative nature of the results is robust to several other values of the discount factor, as discussed in Sec 4.5.

an individual in leisure readership level of *4* gets a one period utility of *1.621* from her being in leisure readership level of *4*. Further, this difference in utility from same level of work and leisure readerships is more pronounced at greater readership levels. Using t-tests we find that the parameters corresponding to work readership are statistically significantly greater than the corresponding parameters for leisure readership.

The results for knowledge state related parameters show that individuals derive higher utility from higher work knowledge than higher leisure knowledge. Specifically, one period utility from being in high work knowledge state is *1.351*. In comparison, the corresponding utility from being in high leisure knowledge state is only *0.116*.

Results for cost parameters show that there are significant costs for posting as well as reading. The cost of only writing one work post is *5.669*. The cost of only writing one leisure post is *4.013*. This shows that the cost of writing a work post is higher than writing a leisure post. We do not find any significant spillover in cost of posting if one posts both leisure and work posts.

There is also significant cost for reading work and leisure posts. Reading only one, two, and three work post costs *2.431*, *2.709*, and *3.792,* respectively. In comparison, reading only one, two, and three leisure post costs *1.703*, *2.260*, and *2.721* respectively. The results show that reading work posts is more costly compared to reading leisure posts. A potential explanation for this is that work-related posts are technically dense and complex and would require significant effort from the reader to internalize[10]. We also find there are significant cost spillover between reading work and leisure posts. For example, the cost of reading one work and one leisure post is *2.431+1.703-0.249*log(2)*. Similarly the cost of reading three work and three leisure posts is *3.792+2.721-0.249*log(10)*. Hence, a reader incurs a lower cost from reading work and leisure posts together than the sum of their individual costs. This finding provides one explanation for the readership spillover effect between work and leisure. One of the major components of the cost of reading is the search cost of finding an appropriate post to read. When an individual reads a work or leisure post of a blogger she is also exposed to her other posts and if she ends up reading two different posts from a blogger her search cost is reduced. In the dataset, we do find a large number of instances where readers read both leisure and work post from the same blogger in a period.

We also find there is significant cost spillover between work reading and posting. The cost of reading 3 work posts and writing one work post is *5.669+3.792-0.113*log(4)*. One potential explanation for this cost

---

[10] In order to test this, we negotiated access to the exact content of the posts from the company and went back to deploy readability analysis of the textual content of the posts. This allows us to check if work-related posts are harder to read in comparison to leisure-related posts. The Flesch Reading Ease Score, which ranges from 0-100 (0 = very hard and 100 = very easy to read) was 45.62 for average work posts and 75.21 for average leisure posts (Kincaid 1975). Thus, the readability analyses support our hypothesis that work-related posts are harder to read than leisure-related posts, and, therefore, the reader would have to spend considerable amount of cognitive effort to understand them. This potentially explains the high cost of reading a work post in our data.

spillover effect is that individuals may get ideas about writing a work post from reading work posts by others. The cost spillover between leisure posting and leisure reading though positive is not significantly different from zero.

5.4. Model Performance

As mentioned earlier, we use the last 8 weeks of data for testing the performance of our model. In this section we present how our model fits the hold out data. We use four test statistics for measuring the performance of our model: (1) distribution of the work posts written, (2) distribution of the leisure posts written, (3) distribution of the work posts read, and (4) distribution of the leisure posts read. We first calculate the values of the test statistics for the test period. We then simulate the posting and reading outcomes for each individual in the holdout period using our model and estimated parameters. We calculate the test statistics for our model by running the simulation 200 times. We compare the distributions obtained from our model with the true distributions in Figure 1. In the figure, the x-axis depicts the test statistic, while y-axis depicts the percentage of individuals corresponding to the test statistic in the holdout sample (on a log scale). The solid blue dots represent the test statistic from the actual dataset, and the boxes-and-whiskers represent the corresponding statistics across the simulated datasets. The whisker represents the upper and lower limits of the 200 corresponding simulated test statistics. The box represents the 25th and the 75th percentile.
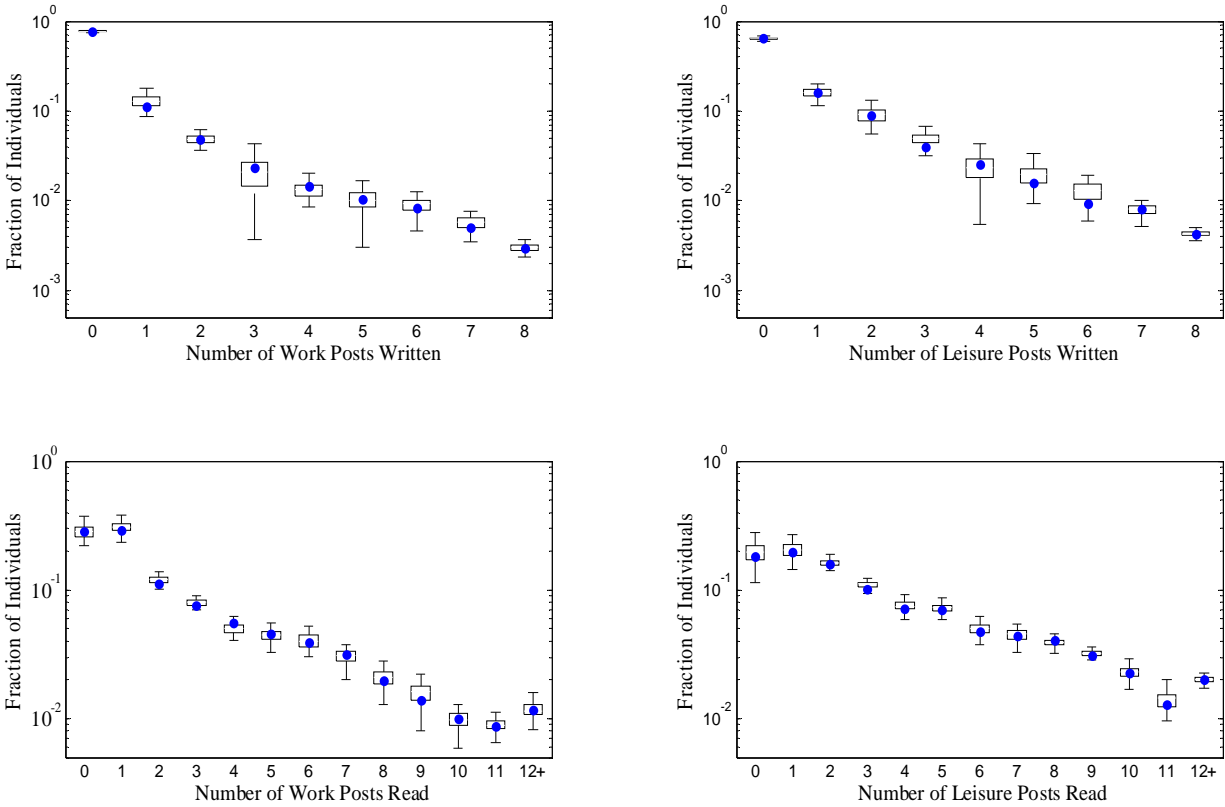
Figure 1: Test Statistics Distribution in Hold-out Data

From Figure 1, we can see that all four test statistics distributions from our model are very close to that for the actual data. Hence, from Figure 1, we can conclude that our model performs very well in predicting key blogging behavior in the holdout sample.

5.5. Robustness Checks

We performed several tests to check the robustness of our results. First, we use alternative measures of readership to test whether the results are robust. We consider three different values for, δ, the readership depreciation factor (0.9, 0.85, and 0.8). When the depreciation factor decreases the past readership depreciates quickly. In all three, cases we find that the qualitatively nature of the results remains the same across these different measure of readership and is consistent with those presented in Table 10. We also considered readership as average readership per post rather than the cumulative readership for each employee. We find that the results from this specification are qualitatively similar to those presented in Table 10.

Second, we test whether individual decision making takes into account peer knowledge states. As mentioned earlier, we have treated knowledge state as an unobserved state. Note, in this robustness check

while peer knowledge state is observable to employees it is not observed by us. This changes step one of our estimation where we estimate the *CCP* and state transition probabilities. We follow the Expectation Maximization (EM) algorithm proposed by Arcidiacono and Miller (2011) to estimate this scenario. This method allows for an individual's decision to be a function of own and peers' observed and unobserved states. The results are statistically similar to the main model results indicating that peer knowledge states do not significantly affect the decision making of individuals.

Third, we have only included one variable for user demographics organizational status. Besides organizational status, unobserved cross-sectional heterogeneity among employees could be a potential driver of the differences in blogging activity. To address this concern, we allow for cross-sectional heterogeneity among employees by allowing their costs of posting and reading to be different. Our results are robust to the inclusion of unobserved individual cross-sectional heterogeneity.

Finally, we conducted a series of additional checks to examine the external validity of the readership and knowledge measures. An external validity for our "knowledge gained from blogging activities" measure could be that knowledge of an employee from reading blogs leads to a higher performance rating for the employee. We collected data on performance evaluations on a subset of employees in our sample. The employee is evaluated by his/her immediate superior and rated on a discrete scale from one to four with one being the lowest and four being the highest performance rating. We find that individuals who are in a higher knowledge state on average receive a higher job performance evaluation.

6. Policy Experiment and Post hoc Analysis

6.1. Policy Experiment

Most enterprises believe that the primary purpose of workplace blogging is to promote work-relevant knowledge sharing. Hence, firms typically like to see employees engage in work-related blogging as opposed to leisure-related blogging. In reality, firms can also implement policies to incentivize employees' blogging activities. In this section, we explore the effect of one such policy intervention: How do employees respond to a policy that prohibits leisure-related blogging?

We introduce the new policy by manipulating the parameter values in the utility function, and then solve for the new equilibrium under the new policy. To execute this experiment, we set the cost of leisure posting and reading to be extremely high. Basically, this policy will eliminate leisure-blogging activities because the costs of leisure posting and reading are always higher than the benefits employees gain from them. One would expect employees to switch from leisure-related activities to work-related activities due to the existence of the budget constraint, and, thus, both work posting and work reading would increase. However, our simulation results suggest a different story. The "No Leisure Policy" column of Table 11

shows the probability of work posting and work reading after the policy is implemented. The probability of work posting decreases from 6.59% to 0.65%. This result is nicely consistent with the readership spillover effect discussed in Section 5.1. When leisure activities are eliminated, employees can no longer benefit from the spillover effect. Hence, the probability of their work posting will drop significantly. As fewer work posts are created the work reading also decreases. Specifically, the work posting decreases by approximately 47% (from 8.04+2*3.38+3*0.83 to 5.05+2*1.65+3*0.29) under then "No-Leisure" policy. This result is consistent with our knowledge state transition results from Section 5.2 which showed that employees are able to learn faster if peers create more posts and more high quality posts are created. As the work posting decreases, the reader's learning from reading slows down. That is the utility they derive from reading a work post goes down. And hence they read even less.

These results have important managerial implications. Firms typically want a larger and more diversified internal knowledge pool. We see that a policy of prohibiting leisure-related blogging can actually hurt the extent of knowledge creation and sharing within the firm. A decrease in the probability of work posting means that fewer people are likely to post and read work-related content. As a result less new knowledge is likely to be added to the knowledge pool.

Table 11: Policy Experiment Results

Mean Probability of Work Reading		
	Current policy	No-Leisure Policy
Read one Work Post	8.04%	5.05%
Read Two Work Posts	3.38%	1.65%
Read Three Work Posts	0.83%	0.29%

Mean Probability of Work Posting	
Current policy	No-Leisure Policy
6.59% (0.02)	0.65% (0.004)

6.2   Post-hoc Analysis - Quantifying the Readership Spillover Effect

Using the structural parameter estimates, we now perform analysis that quantifies the effect of spillover from leisure readership to work readership and vice versa on amount of leisure and work postings. To identify the effect or readership spillovers on blogging behavior, we modify the parameter corresponding to readership spillover in the readership state transitions probabilities to zero. We then solve for the equilibrium. We consider two different state transition structures (1) no spillover from leisure readership to work readership, (2) no spillover from work readership to leisure readership. In both scenarios, we also set the reading cost spillover parameter for leisure and work to zero. The results are reported in Table 12.

Table 12: Effects of Readership Spillover on Blogging behavior

Equilibrium Probabilities	(0) No Change	(1) No Readership Spillover from Leisure to Work	(2) No Readership Spillover from Work to Leisure
Write Work Post	6.59%	0.62%	2.17%
Write Leisure Post	10.43%	6.94%	7.95%
Read one Work Post	8.04%	6.05%	6.26%
Read Two Work Posts	3.38%	2.53%	2.48%
Read Three Work Posts	0.83%	0.62%	0.72%
Read One Leisure Post	12.41%	10.25%	11.08%
Read Two Leisure Posts	3.69%	3.03%	3.39%
Read Three Leisure Posts	1.55%	0.84%	0.85%

For comparison purposes, we also report the equilibrium probabilities for posting and reading under the real scenario (no change in readership spillover). From the results of model (1) and model (0), we can see that if there were no spillover from leisure readership to work, the work posting would decrease by 90.59% (from 6.59% to 0.62%) and the leisure posting would decrease by 33.46% (from 10.43% to 6.94%). From these results we can attribute 33.46% of leisure posting to the spillover effect from leisure to work readership. Similarly, 90.59% of work posting is due to the leisure to work readership spillover effect. It is easy to see how both work and leisure posting decrease if the spillover from leisure to work readership is removed. In the presence of spillover from leisure to work readership, leisure provides a direct utility (from cumulative leisure readership) and indirect utility (by helping increasing the cumulative work readership). Hence, employees derive higher utility from writing a leisure post in the presence of spillover effect than otherwise. Similarly, employees derive higher utility from writing work post (due to higher cumulative work readership) compared to what they would derive in the absence of the spillover.

We can see if there were no readership spillover from work to leisure then work (leisure) posting would decrease by 67.07% (23.78%). These results indicate that the spillover from leisure to work readership has a larger impact on individual blogging behavior than spillover from work to leisure. Also note that in all cases, both work and leisure reading would also decrease if there were no readership spillovers. The reading goes down because, in the absence of readership spillovers readers have less content to read and gain knowledge from. These results highlight the interdependence between leisure and work posting and between posting and reading.

6.3. Post-hoc Analysis - Quantifying the Competition Effect

We now present analysis that helps us identify the effect of competition for readership on blogging behavior

of individuals. To change the level of competition we can artificially reduce or increase the number of peers who write posts. However, the number of peers who post per period is endogenously determined. So artificially reducing or increasing the number of peers who post would make the model internally inconsistent. Hence, for identifying the competition effect we modify the parameters corresponding to the peer related variables (other than the number of posts read by peers) used for calculating cumulative readership state transitions (Equation 9) and then solve for equilibrium. We consider two levels of competition and report the results in Table 13. We report results for the model where (1) the parameters corresponding to the competition are reduced by 10%, and (2) the competition specific parameters are increased by 10%. We can see that if the effect of competition on state transitions decreases by 10% the equilibrium mean work posting and leisure posting probabilities increase by 18.06% and 14.77% respectively. Similarly a 10% increase in effect of competition on state transitions decreases equilibrium mean work and leisure posting by 11.34% and 13.36% respectively. Further, both work and leisure reading decrease as competition increases because less content gets generated from which they can gain knowledge.

Table 13: Effects of Competition on Blogging behavior

Equilibrium Probabilities	(0) No Change	(1) 10% Lower Competition for Readership	(2) 10% Higher Competition for Readership
Write Work Post	6.59%	7.78%	5.61%
Write Leisure Post	10.43%	11.97%	8.83%
Read one Work Post	8.04%	8.95%	7.12%
Read Two Work Posts	3.38%	3.82%	3.07%
Read Three Work Posts	0.83%	0.96%	0.69%
Read One Leisure Post	12.41%	13.44%	11.31%
Read Two Leisure Posts	3.69%	3.96%	3.04%
Read Three Leisure Posts	1.55%	1.63%	1.27%

7. Conclusion

Over the past few years, we have seen that organizations have deployed enterprise social-media technologies widely for internal use. Companies can use blogs internally to open conversations between management and employees. If employees feel that they are part of the conversation, one might see more loyalty and better productivity. A blog is also useful for product development and for educating the company about upcoming initiatives. Blogs can be more effective than email since they eliminate the need to answer the same question more than once and allow for feedback from the whole group as opposed to just the email recipient. Blogs allow firms to trace employees' personal expertise and practices. Making these visible provides an idea of who knows what, which is a starting point for collaboration and for allowing knowledge to spread more

effectively.

In this paper, we present a dynamic structural model in which employees in an enterprise compete with each other in the process of reading and writing blog posts. There are two kinds of blogs posts in our context: work-related and leisure-related posts. Users make choices about reading and writing based on their preferences for either type of content. Our findings suggest that individuals derive a positive utility from increases in the cumulative readership of their work- and leisure-related posts. Work-related blogging allows employees to express their expertise, and once they are identified as "experts," this reputational gain can lead to economic benefits. In addition, leisure readership can indicate how popular employees are among their professional peers, and the more popular they are, the happier they are likely to feel at work. However, given the same amount of cumulative readership for work and leisure, employees derive 27% to 71% higher utility from work-related cumulative readership as compared to utility from leisure-related cumulative readership. Further, reading and writing work-related posts is more costly than reading and writing leisure-related posts, on average.

We find evidence of competition among employees with regard to attracting readership for their posts. We identify a tension that peer blogging activity raises for an employee. As a blogger, the employees would prefer lower competition for readership and, hence, would prefer fewer of her peers to post. In contrast, as a reader the same employee would prefer peers to post more as that increases her rate of learning.

While readership of leisure posts provides less direct utility than that of work posts, employees still post a significant number of leisure posts, as there is a significant spillover effect on the readership of work posts from the creation of leisure posts. Because the two types of blog posts co-exist on the same platform, this spillover is bidirectional in nature. Our policy simulations suggest that prohibiting leisure-related posting will be counter-productive for organizations since it also leads to a reduction in work-related posting as well as reading. Overall, these results shed light on how enterprise adoption of social-media tools influences employee blogging behavior and choices.

An interesting implication of one of our robustness checks that look at external validity of cumulative readership and knowledge gained from blogging is that work-performance rating is positively correlated with the knowledge states of employees. This indicates that employees who have gained higher levels of knowledge from their blogging activities tend to perform better in the organization. While this correlation is merely suggestive, it is, nonetheless, a useful insight because many organizations have been concerned about the effects of blogging on employee performance and productivity.

Our paper has several limitations. First, in terms of reading, we model only whether an individual

reads work-related or leisure-related posts. We do not model whose blog an individual reads.[11] Individual reading dynamics may be affected by the dynamics of content created by the bloggers they follow. Future research can explicitly model this relationship. Second, commenting on blogs can be an important part of the knowledge-creation process, but we do not have sufficient information on the comments on the blog posts in our data to do any meaningful analyses. Third, due to data limitations, we account for blog- reading behavior only within the enterprise. Some of the blog creation and consumption dynamics may be affected by events outside of our data context. For example, employees can also access external blogs and other social-media content. However, such data are not available to us. Notwithstanding these limitations, we hope our work can pave the way for future research in this emerging area of enterprise social media.

References

Aggarwal, R., Gopal, R., Sankaranarayanan, R., and Singh, P. V. 2011. Blog, Blogger and the Firm: Can Negative Posts by Employees Lead to Positive Outcomes? *Information Systems Research* forthcoming.

Aguirregabiria, V., and Mira, P. 2007. Sequential Estimation of Dynamic Discrete Games. *Econometrica (75:1)*, pp 1-53.

Ahn, D., Duan, J., and C. Mela. 2011. An Equilibrium Model of User Generated Content, Working Paper SSRN.

Albuquerque, P., Pavlidis, P., Chatow, U., Chen, K., Jamal, Z., and Koh, K. 2012. Evaluating Promotional Activities in an Online Two-Sided Market of User-Generated Content. forthcoming, Marketing Science.

Arcidiacono, P., and Miller, R. 2011. CCP Estimation of Dynamic Models with Unobserved Heterogeneity. *Econometrica (76:6)*, 1823-1867.

Argote, L., McEvily, B., and Reagans, R. 2003. Managing Knowledge in Organizations: An Integrative Framework and Review of Emerging Themes, *Management Science (49:4)*, 571-582.

Bajari, P., Benkard, C., and Levin, J. 2007. Estimating Dynamic Models of Imperfect Competition. *Econometrica (75:5)*, 1331-1370.

Bajari, P., Chernozhukov, V., Hong, H., and Nekipelov, D. 2009. Nonparametric and Semiparametric Analysis of a Dynamic Discrete Game. Unpublished manuscript, Dep. Econ., Univ. Minn., Minneapolis.

Breusch T. S. and Pagan A. R. 1980. The Lagrange Multiplier Test and its Applications to Model Specification in Econometric, *The Review of Economic Studies(47:1)*, 239-253.

Ching, A. 2010. A Dynamic Oligopoly Structural Model for the Prescription Drug Market After Patent Expiration. *International Economic Review, (51:4)*, pp.1175-1207.

Chung, D., Steenburgh, T., and K. Sudhir 2009. Do Bonuses Enhance Sales Productivity? A Dynamic Structural Analysis of Bonus Based Compensation Plans, Working Paper.

Cohen,W., and D. Levinthal. 1990. Absorptive capacity: A new perspective on learning and innovation", *Administrative Science Quarterly*, *(35:1)*, 128-152

Constant, D., Kiesler, S., and Sproull, L. 1994. What's Mine Is Ours, or Is It? A Study of Attitudes About Information Sharing. *Information Systems Research (5:4)*, 400-421.

---

[11]There are multiple issues that make estimation of a structural model with dyadic readership intractable. First, the state space would blow up. Second, the choice space would also blow up.

Dewan, S., and J. Ramaprasad 2012. Music Blogging, Online Sampling, and the Long Tail. 2012., Information Systems Research, Article in advance.

Dubé, J. P., Sudhir, K., Ching, A., Crawford, G., Draganska, M., Fox, J., Hartmann, W., Hitsch, G., Viard, B., Villas-Boas, M., Vilcassim, N. 2005. Recent Advances in Structural Econometric Modeling: Dynamics, Product Positioning and Entry. *Marketing Letters*, *(16:3-4)*, 209-224.

Erdem, T., Keane, M., and Sun, B. 2008. The Impact of Advertising on Consumer Price Sensitivity in Experience Goods Markets. *Quantitative Marketing and Economics (6:2)*, 139-176.

Ericson, R., and Pakes, A. 1995. Markov-Perfect Industry Dynamics: A Framework for Empirical Work. *The Review of Economic Studies (62:1)*, 53-82.

Farias, V,. D. Saure, and G.Y. Weintraub. 2012, An Approximate Dynamic Programming Approach to Solving Dynamic Oligopoly Models, *The RAND Journal of Economics*, *(43:2)*, 253-282.

French, J. R. P., and Raven, B. H. 1959. The bases of social power. In D. Cartwright (Ed.), Studies in social power, Ann Arbor, MI: Institute for Social Research, 150-167.

Ghose, A., and Han, S. 2011. An Empirical Analysis of User Content Generation and Usage Behavior on the Mobile Internet, *Management Science,* 57(9), 1671-1691.

Ghose, A., and Han, S. 2010. A Dynamic Structural Model of User Learning in Mobile Media Content. Working Paper, SSRN.

Hofstetter, R., Shriver, S., and Nair, H. 2009. Social Ties and User Generated Content: Evidence from an Online Social Network. Working Paper, Stanford University.

Hotz, V., and Miller, R. 1993. Conditional Choice Probabilities and the Estimation of Dynamic Models. *The Review of Economic Studies (60:3)*, 497-529.

Hausman, J. A. 1978. Specification Tests in Econometrics, *Econometrica(46:6),* 1251-1271.

Huh, J., Jones, L., Erickson, T., Kellogg, W., Bellamy, R., and Thomas, J. 2007. Blogcentral: The Role of Internal Blogs at Work. *Conference on Human Factors in Computing Systems*, 2447-2452.

Kavanaugh, A., Zin, T., Carroll, J., Schmitz, J., Pérez-Quiñones, M., and Isenhour, P. 2006. When Opinion Leaders Blog: New Forms of Citizen Interaction. *Proceedings of the 2006 international conference on Digital government research*, 79-88.

Kumar. V., and Sun, B. 2010. Why do Consumers Contribute to Connected Goods? A Dynamic Game of Competition and Cooperation in Social Networks. Working Paper.

Lee, S., Hwang, T., and Lee, H. 2006. Corporate Blogging Strategies of the Fortune 500 Companies. *Management Decision (44:3)*, 316-334.

Levy, M. 2009. WEB 2.0 implications on knowledge management, *Journal of Knowledge Management (13:1)*, 120-134.

Lu, Y., Singh, P. V, and Sun, B. 2010. Blind Men Can Judge No Color: A Dynamic Structural Analysis of Enterprise Knowledge sharing. Working Paper.

Lucas, R. 1976. Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, *(1:1)*, 19–46

McAfee, A. P. 2006. Enterprise 2.0: The Dawn of Emergent Collaboration. *MIT Sloan Management Review, (47:3)*, 21–28.

McAfee, A. P. 2009. Enterprise 2.0 is vital for business, *Financial Times (9 December)*.

Nardi, B., Schiano, D., Gumbrecht, M., and Swartz, L. 2004. Why We Blog. *Communications of the ACM (47:12)*, 41-46.

Pakes, A. and Mcguire, P. 1994. Computing Markov-Perfect Nash Equilibrium: Empirical Implicaitons Of A Dynamic Model. *Rand Journal of Economics*, 555-589.

Pesendorfer, M., Schmidt-Dengler, P., and Street, H. 2003. Identification and Estimation of Dynamic Games. NBER working paper.

Rabiner, L. R. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE (77),* 257-286.

Roberts, J. A, Hann, I. H, Slaughter, S. A. 2006. Understanding the Motivations, Participation, and Performance of OSS Developers: A Longitudinal Study of the Apache Projects, *Management Science (52:7)*, 984-999.

Singh, P., Sahoo, N., and Mukhopadhyay, T. 2010a. Seeking Variety: A Dynamic Model of Employee Blog Reading Behavior. Working Paper.

Singh, P.V., Tan, Y., and Youn, N. 2010b. A Hidden Markov Model of Developer Learning Dynamics in Open Source Software Projects. *Information Systems Research,* forthcoming.

Sun, M., and F. Zhu. 2012. Ad Revenue and Content Commercialization: Evidence from Blog, Working paper SSRN.

Thomas-Hunt, M. C., Ogden, T. Y., and Neale, M. A. 2003. Who's really sharing? Effects of social and expert status on knowledge exchange within groups. *Management Science (49:4)*, 464-477.

Trusov, M., A. Bodapati and R. E. Bucklin. 2010. Determining Influential Users in Internet Social Networks. *Journal of Marketing Research, (47:4)*, 643-658.

Wattal, S., Racherla, P., Mandviwalla, M. Network Externalities and Technology Use: A Quantitative Analysis of Intraorganizational Blogs. *Journal of Management Information Systems, (27:1)*, 145-174.

Weintraub, G.Y., C.L. Benkard and B. Van Roy (2008), Markov Perfect Industry Dynamics with Many Firms, *Econometrica, (76:6)*, 1375–1411.

Yardi, S., Golder, S., and Brzozowski, M. 2009. Blogging at Work and the Corporate Attention Economy. *International Conference on Human factors in Computing Systems,* 2071-2080.