Pre-analysis plan

Media and Motivation: The effect of performance pay on writers and content

Jared Gars and Emilia Tjernström
University of Wisconsin, Madison

May 15, 2017

1 Introduction

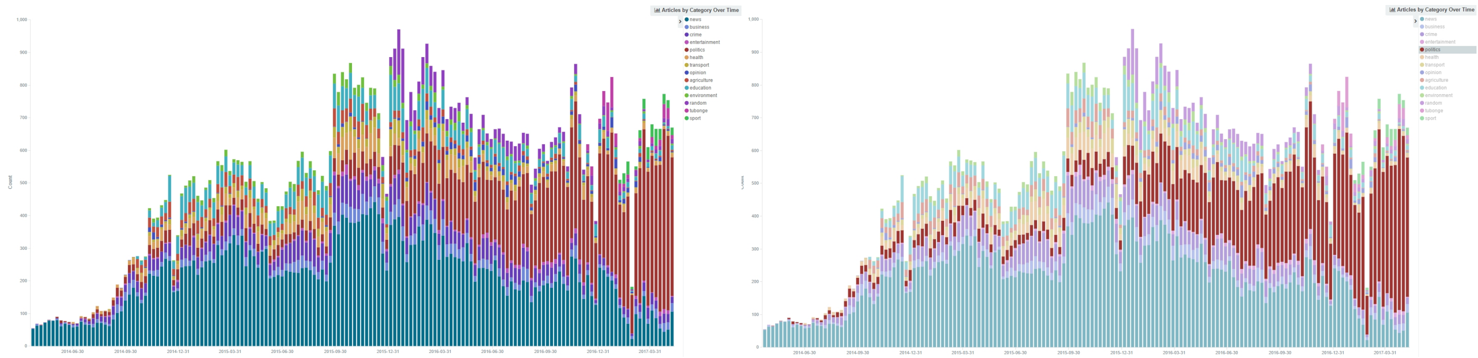
Performance contracts have been implemented in a variety of settings (Lazear, 2000; Finan et al., 2015) yet there is mixed empirical evidence on whether performance pay actually matters for overall firm profits (Prendergast, 2015).[1] When the marginal productivity of effort is random, learning from performance feedback may be costly and the added risk forces firms to provide insurance to compensate workers (Prendergast, 2002). Output-based contracts may be more efficient than input-based contracts when agents have specific knowledge about the production function (Raith, 2008), or can learn the returns to specific actions. Further, contracting on outcomes may crowd out intrinsic motivation (Bénabou and Tirole, 2006), incentivize short-termist behavior that rewards only current output but may erode long term productivity (Bolton et al., 2006), and lead to a reduction in effort allocated to non-contracted outcomes (Holmstrom and Milgrom, 1991).

There has recently been a shift towards performance based contracts for online journalists (also known as "pay-per-click" contracts) according to a report by the Columbia Journalism Review (2016). Despite concerns about the impacts of this business model on the quality of journalism[2], there exist no studies that investigate the relationship between performance contracts and firm profits or journalistic quality in the market for news. In this project, we will use an experimental design to study the impacts of pay-per-click contracts for journalists on writer performance, firm profits and journalistic quality in an online news firm in Kenya.

We seek to provide evidence on the tradeoffs between the strength of incentives and the effectiveness of contracting on outputs in a context where agents have (or can gain) specific

[1] See Bandiera et al. (2011) or Miller and Babiarz (2013) for recent overviews of empirical evidence on pay-for-performance contracting in developing countries.

[2] Concerns include the risk that journalists choose to dumb down stories, increasing their use of easy topics and 'click-bait' headlines. Supporters of the model argue that readers value investigative journalism, which will be rewarded when journalists have greater discretion over how to connect with their readers.



(a) All topics (b) Politics highlighted

Figure 1: Number of articles published over time, by topic

knowledge about the returns to various actions. In so doing, our research design will focus on two main contributions:

First, combining random allocation to contract type with individual-level productivity beliefs, we will test whether stronger incentives increase the acquisition and use of information on marginal returns to writer effort. While some theoretical work focuses on contracting when the marginal productivity of effort is stochastic (Zabojnik, 1996; Rantakari et al., 2008), empirical tests separating the information effect from the incentive effect of performance contracts remain limited.

Second, by linking writer-level ethnicity with crowdsourced measures of article-level ethnic and political bias, we will study the relative importance of supply- and demand-side sources of media bias across contract type. Theories of supply side sources of media bias suggest a role for wages and journalistic discretion on the incentives for reporters to bias news to promote their worldview (Baron, 2006), while demand side theories argue that bias only occur in equilibrium if firms find it profitable to distort information to match consumers' ideologies (Mullainathan et al., 2005; Gentzkow and Shapiro, 2006). We will test the degree to which bias in reporting is driven by writers' ethnic ties or the distribution of political preferences of consumers.

2 The intervention

2.1 The firm

The study will be conducted using within-firm data from a digital news platform in Kenya. The website is an online, local, county-level Kenyan newspaper tailored for the mobile web. It sources all its news stories from local reporters who are currently paid via mobile-money for each article they publish. Articles are tagged by topics, such as politics, business and agriculture, with politics being the most prominent topic. Writers do not need previous qualifications to submit content, only a device and basic writing skills.

After signing up on the webpage, citizen reporters submit articles, subject to a 500-word limit. Professional editors curate the incoming articles and choose which ones to publish, monitoring for defamation, plagiarism, hate speech, and a few more issues that guarantee rejection. If a citizen reporter's article is published, the reporter is paid 100 Ksh via the mobile money system M-Pesa. The paper receives over a million monthly views as of late 2016, and this number is steadily growing: average daily views grew from 17,000 in January 2016 to about 137,700 in December, 2016. Articles are written by more than 1,000 writers, who publish over 4 articles per week.

2.2 Treatments

The project aims to examine the effects of different types of incentives for writers on the quality and quantity of their output. We will compare three different types of contracts:

- *Control group:* The control group contracts will remain the same. The current reward structure is a fixed fee of w_a Ksh per article that is published. The incentives for writers are therefore to pass the basic editorial bar (which consists of no plagiarism, no hate speech, no defamation), but little incentive to improve article quality beyond that basic threshold.

- *Treatment group A / Pay-per-view:* This contract will reward writers with a fee structure composed by a smaller fixed fee, and a bonus for every click received. In other words, $w = w_b + b * views$, where b will be calibrated using the prior month's views, such that the ex ante expected value returns for writers equal to the status quo of w_a Ksh per article.

- *Treatment group B / Contract choice:* The third group will be allowed to select at the beginning of the treatment period whether they would like to stay with the pay-per-article contract, or if they would like to switch to a pay-per-view system. If they choose the pay-per-view contract, the contract will be identical to Treatment group A.

Contract parameters and design

The contract parameters will be determined by equalizing the expected returns per article to each contract for the median writer. Given the stochastic nature of output based pay in the market for news, more risk averse agents may need to be compensated for increased risk, and therefore expected compensation must be greater all else equal. However, empirical evidence on the incentives-risk tradeoff remains scarce, and the relation between risk and performance contracting may be "too subtle" to consider in the design of actual compensation plans (Raith, 2008). Various factors to take into account when choosing the values of the parameters include: journalists' career concerns and preference for influence, specific knowledge of the production function, incentives to learn.

2.3 Sample description

We will define our sample as consisting of the active writers at the time when the study begins, i.e. those that have written at least one article in the last month. There is a lot of variability in the number of articles that writers submit per week, and we want to explore treatment heterogeneity across different levels of experience within the sample, so we set this bar fairly low. Currently, there are around 1,000 active writers on the site, and the average writer publishes four articles per week.

2.4 Allocation to treatment

Writers will be randomly allocated to treatment groups, but we will stratify the randomization on average views per week and on writer ethnicity. We also have data on the total number of past articles that each writer in the sample have written, as well as how long they have been signed up to the website, and their ratio of submitted articles to published articles, but average views and ethnicity seem to be the most important things to stratify on.

2.5 Power calculations

We have little ex ante information on how writers might respond to this type of incentive, but we have information on the mean number of articles that writers submit. Power calculations suggest that our minimal detectable effect would be around 0.3 of a standard deviation for our worst-case sample size and closer to 0.2 of a standard deviation if we manage to recruit the sample size that we hope to be able to get.[3] In terms of the other outcome measures, we have insufficient data at the moment to base power calculations on. Since we have not yet completely defined our measures of bias, and thus have not yet begun coding articles accordingly, we are not certain to be able to conduct power calculations on these outcome measures.

3 Data

3.1 Writer-level output data

Using the firm's existing digital submission framework, in conjunction with data from Google analytics, we will create writer- and article-level measures of the number of articles submitted,

[3] There are still some uncertainties about how many writers will qualify, since the company recently transferred to a new data management system, meaning that we only have about 2 months worth of data on these particular variables. Basic power calculations were done in Optimal Design, setting $\alpha = 0.05$ and Power = 0.8, and assuming that we have a covariate that explains roughly 30% of the variation in the outcome variable, which seems to be the amount of explanatory power that we will get with pre-treatment average views. With a mean number of articles per week of $\mu = 4.2$, with a standard deviation of around 6, 0.3 of a standard deviation would translate into being able to statistically detect impacts of an increase of 1.8 articles/week for the smaller sample size and 1.2 articles per week with the larger sample size.

views, shares on social media, average earnings, and composition of articles submitted (topics are pre-defined by existing article categories: politics, sports, entertainment, etc).

3.2 Subjective beliefs

Prior to submitting each article, writers will be asked to guess *(i)* whether the article will pass the editor, and *(ii)* to predict the number of views that the article will receive in the following week. Writers will respond by typing the answers to the following questions: First, they will answer "What is the likelihood that this article will be accepted for publication?"[4], which provides an estimate of the writer's subjective beliefs about the minimum quality of the article. Second, they will answer "About how many views do you expect this article to receive in the following week?", which will provide us with an estimate of the writer's subjective mean about the 'productivity' of the article. For a subset of article submissions, we will incentivize correct predictions to compare with the larger sample of non-incentivized beliefs.

3.3 Risk aversion

Since pay-per-click constitutes a stochastic contract, we may see differential impacts depending on writers' level of risk aversion. Firm profits may also vary depending on whether writer satisfaction with the contracts requires greater compensation for increased risk, so we will measure risk aversion among the writers to explore heterogeneous impacts.

3.4 Article bias

For each article, we will enlist auditors to carry out the following tasks: First, they will assign the articles into primary article classifications (politics, business, etc) based on a list of up to 10 potential classifications.[5] Second, the article will be classified as news or opinion. Third, auditors will be asked to state the political parties mentioned in the article. For each party that is mentioned, auditors will answer the question "Is this article generally positive, neutral, or negative towards this political party?" Answers will be given on a 5-point scale, from very positive to very negative, including an option of "don't know/not applicable".

We will take the article-level score, normalize it to [-1,1] and average it across auditors (the number of auditors per article is still TBD), which will provide one or several party-favorability scores for each article in the politics section. Of course, the overall negativity of an article could either be due to the writer's ideological slant or to the true nature of the event/topic. However, since writers choose what topics to write about, a writer who chooses to write about topics that

[4]Since this is a relatively well-educated population, we think we will be able to elicit simple probabilities. We will pilot this, and may change it to a Likert scale if writers struggle with the probabilities

[5]The current topics are Politics, Opinion, Business, Agriculture, Crime, Sports, Entertainment, Education, Environment, and Transportation. The articles will already be classified on the website, but we want to verify that this categorization is correct.

are consistently negative about a specific ethnicity/party is likely biased. Since negative news or events are time-specific, we will define writers as biased if their articles are consistently more negative about a party than the average article.Specifically, we will define a party-bias dummy variable, $[party]\_bias_{it}$, which equals 1 if writer i's average article is more negative about [party] than the average article in time t (time periods will be weeks).

Using this approach, we will be able to assess writer slant, which in this case encompasses both the issues that writers choose to cover, and for 'bias-prone' topics, whether their coverage is perceived as fair or slanted. Auditors will be Kenyan Mechanical Turk members and/or Kenyan students, who will be screened based on their capacity to recognize journalistic quality, their political knowledge, and their capacity to categorize the content of some subset of articles.

Measures of writer ethnicity: We hope to code writers' ethnicities, to judge whether their topics and/or bias varies by the ethnicity of the actors in the article. We will attempt a classification that builds on (Harris, 2015). Harris collects data on names from ethnically homogeneous regions of Kenya, and builds a training data set for these ethnic groups. This can then be used to predict writers' ethnicities based on their names. We will combine this information with data gleaned from writers' IP address, which places them in a county, for which we can obtain estimates of ethnic composition.[6] We may alternatively elicit guesses about writer ethnicity from Kenyan MTurk members for all the writers in our sample; preliminary discussions with Kenyan nationals indicates that ethnicity can be accurately guessed for the majority of Kenyans. If we go this route, we will elicit ethnicities from at least 3 auditors per writer, and evaluate the certainty of the guess based on whether all auditors select the correct ethnicity for a given writer.

If we manage to obtain good measures of ethnicity, we will then be able to investigate whether co-ethnicity influences the bias of articles, and whether this effect varies by type of contract. One issue related to ethnicity is that measurement error in the ethnicity variable will likely be non-classical, since we will have less training data/more inaccuracy for smaller ethnicities. Also, in many areas of northeastern Kenya, which are predominantly Muslim, people are increasingly adoption Arabic or Islamic names, making ethnic categorization difficult. Both readership and writers from those regions are minimal, so it should not be a major concern.

4 Hypotheses

We aim to test 5 sets of hypotheses. The first set constitute the main average treatment effects, which relate to the effect of contract type on writer behavior and performance. Beyond estimating the causal effect of contract type, the other sets of hypotheses investigate different mechanisms, to shed light on the margins over which writers change their behavior, the implications for ideological slant, the degree to which the pay-per-view contract encourages writers

[6] We will have writers' names, since they report them when they start to work for the media firm.

to innovate, and the role of selection on the treatment effects.

- **Hypothesis 1**: Writers will alter their effort allocation in response to the pay-per-view incentive, along the dimensions of effort and/or article characteristics. Specifically, we will examine the following effects:
  - **1 a)** If writers have knowledge about the production function that relates article-level characteristics to views, then we expect the average views per week for writers under PPV to increase relative to the control group over time, with little impact on the output variance.
  - **1 b)** If writers initially do not know the production function, but can learn about it by experimenting, then we expect the variance of average views per week under PPV to increase in the beginning as writers experiment, followed by an increase in the average views per week under PPV relative to the control group.
  - **1 c)** If writers are unable to learn about the production function, we may find effects of the treatment on the number of articles submitted per week, and/or on other article characteristics, such as topic.
  - **1 d)** We also expect the effects of PPV to be heterogeneous along several writer-level dimensions – in particular baseline experience, risk aversion, baseline productivity beliefs, and ethnicity.
- **Hypothesis 2**: The average writer under PPV contracts will alter their choice of topics in response to the contract. Specifically, we will examine the following effects:
  - **2 a)** The pay-per-view contract may skew the composition of topics for an individual writer towards high-demand topics, as defined by high average views.
  - **2 b)** Risk-averse writers will prefer low-variance topics, rather than high-demand topics.
- **Hypothesis 3:** The contract structure may affect the bias in writers' articles. Assuming that readers have biased beliefs related to ethnicity and party affiliation, and that they value seeing these beliefs confirmed (as in Mullainathan and Shleifer, 2005), we would expect the ethnic bias of writers in PPV contracts to skew towards the ethnic composition of the readership.
- **Hypothesis 4:** We expect learning about performance to differ between the different contracts.
  - **4 a)** Specifically, we expect writers in Treatment group A (pay-per-view) to improve their productivity predictions at a faster rate than writers in the control group.

- **4 b)** As described in H1b, we expect that the variance of weekly views may increase during a period following treatment, in which writers engage in a process of "learning by doing" that may vary with writer experience and their risk tolerance.

- **Hypothesis 5:** We will also explore the selection mechanisms, both in terms of who selects into what contract among Treatment group B (contract choice) and in terms of who (if anyone) quits in Treatment group A:
  - **5 a)** We expect take-up of pay-per view in the contract choice group to vary with risk aversion, with more risk averse writers being less likely to take up the new contract.
  - **5 b)** We expect writers with above-average views pre-treatment to be more likely to take up the pay-per-view contract.
  - **5 c)** We expect overconfident writers at baseline to be more likely to take up the pay-per-view contract.[7]

5 Methodology

5.1 Balance tests

We will test for balance on a battery of pre-treatment variables between the three groups (control + two treatment groups). Specifically, we will examine pairwise t-tests of the null hypothesis of equality of means between each treatment group and the control group, as well as Kolmogorov-Smirnov tests of equality of distribution for continuous variables. We will also conduct a joint test of orthogonality by regressing a dummy for treatment on these same covariates, and testing the hypothesis that all the coefficients on the covariates are jointly equal to zero with an F-test.

The covariates that we will test include (we may add more):

- Number of articles published week & average earnings[8]
- Total number of past articles
- Tenure (weeks since signed up)
- Average views per article
- Ratio of submitted to published articles
- Share of articles on each of 10 topics
- Average bias

[7] We will begin to ask writers about their beliefs before treatment is assigned, so we will have baseline data on this variable.

[8] Before the treatment, average earnings will be perfectly determined by the number of articles published.

- Ethnicity
- Productivity beliefs (expected views)

5.2 Estimation strategy

First, we estimate the average impact of assignment to the pay-per-view contract on writer behavior and productivity. In the absence of input observability, rewarding outputs may be more efficient when agents have specific knowledge of the production function (or can learn) and that these benefits outweigh the cost of compensating agents for increased risk. In the absence of learning, treatment writers should shift their effort or article characteristics to increase their revenues. We test whether the pay-per-click contract changes writers' behavior on a number of dimensions. Following McKenzie (2012), we pool the data and estimate the following ANCOVA model to estimate the average treatment effects, controlling for the baseline value of the outcome variables y_{i0}, other baseline individual characteristics X_i, strata fixed effects π_s, and time fixed effects λ_t:

$$y_{it} = \alpha + \beta_1 Treat_i^j + \gamma y_{i0} + \eta X_i + \pi_s + \lambda_t + \epsilon_{it} \tag{1}$$

Treatments are indexed by j and include both the pay-per-view treatment and the contract selection treatment. Outcome variables of interest at the week level include: the number of articles published, total views, earnings, the likelihood of quitting, and the number of views per article published. Baseline values of the outcome variables are calculated by averaging the weekly values over the 6 weeks prior to the introduction of the treatment. Writer-level characteristics include risk preferences, experience (both tenure and total number of published articles), ethnicity, and county-level location. We will cluster our standard errors at the strata-level.

While the specifications described by Eq. 1 provide average treatment effects across workers by contract type, they do not pin down changes in within-worker productivity across contract types. We will therefore also estimate the average treatment effect by pooling together weekly writer-level measures before and after the treatment and estimate the following regression, with and without writer level fixed effects:

$$y_{it} = \alpha + \beta_1 Treat_i^j + \lambda_t + \pi_s + \theta_i + \epsilon_{it} \tag{2}$$

where t includes weekly worker measures in the 6 weeks prior to the introduction of the treatment and θ_i are individual fixed effects .

To explore whether the effects of pay-per-view contracts are heterogenous across writer characteristics, we interact treatment with multiple writer level measures that may affect their responsiveness to the performance contracts including risk preferences (Prendergast, 2002), experience, baseline productivity beliefs, and ethnicity. We modify the standard ANCOVA estimation to allow the treatment effects to vary by these characteristics (C_i^j):

$$y_{it} = \alpha + \beta_1 Treat_i^j * C_i^k + \beta_2 Treat_i + \beta_3 C_i^k + \theta y_{i0} + \pi_s + \lambda_t + \epsilon_{it} \tag{3}$$

After estimating the causal impacts of the contracts, we will investigate a number of mechanisms to determine the margins over which writers change their behavior, the implications for ideological slant, the degree to which the pay-per-view contract encourages writers to innovate, and the role of selection on the treatment effects.

Content

First, we test whether the contract type determines the content in two ways: topic choice and bias. To test the former (Hypothesis 2), we will rank topics (currently defined across broad groups that include news, politics, crime, entertainment, etc.) and construct a writer-week measure of the weighted share of articles by the topic rank of average pageviews at baseline. For example, if politics has the highest average rank (10) and agriculture has the lowest (1), a writer that submits two agriculture articles and one politics article in a week will have a value: $share\_avgviewsit = .33*10+.66*1 = 3.96$. We will estimate the following ANCOVA regression to test the impacts of contract type on the composition of article types:

$$share\_avgviews_{it} = \alpha + \beta_1 Treat_i^j + \gamma y_{i0} + \eta X_i + \pi_s + \lambda_t + \epsilon_{it} \tag{4}$$

In addition to average views, more risk averse agents should prefer topics with lower variance of pageviews under a pay-per-view contract. We construct a ranking of topics by their variance and calculate a measure of writer-week article share by the relative ranking of pageview variance, $share\_varviews_{it}$ and interact the treatment with writer risk preferences. We will run the following modified ANCOVA regression:

$$share\_varviews_{it} = \alpha + \beta_1 Treat_i^j * CRRA + \beta_2 Treat_i + \beta_3 CRRA + \theta y_{i0} + \pi_s + \lambda_t + \epsilon_{it} \tag{5}$$

and test whether writer risk aversion determines topic selection differentially in the pay-per-view contract, when risk is introduced into the writer's compensation.

Bias

Using the crowdsourced bias measures described in Section 3.4, we will test whether the pay-per-view contract changes the amount of ethnic or political slant produced by writers. First, we test whether average bias is higher in the treatment group by estimating the following regression:

$$I_{it} = \alpha + \beta_1 Treat_i^j + \gamma I_{i0} + \eta X_i + \pi_s + \lambda_t + \epsilon_{it} \tag{6}$$

where I_{it} is a weekly average measure of writer $i's$ ideological slant and I_{i0} is the baseline value calculated over the prior 6 weeks. We will estimate this regression separately using both the bias measure of article content and the title to test if writers in the treatment produce more “click-bait" titles and/or alter the composition of their articles.

Conditional on readers having biased beliefs related to ethnicity and party affiliation, and that they value seeing these beliefs confirmed (as in Mullainathan and Shleifer, 2005), we would expect the ethnic bias of writers in the pay-per-view treatment to skew towards the ethnic composition of the readership.

Learning

In contrast to input-based performance pay, rewarding outputs may strengthen the incentives for writers to exploit their contextual knowledge, use their existing inputs more efficiently, or innovate through experimentation in writing topic, quality, or bias, if the expected benefits exceed the increased risk. To identify writer learning or innovation, we first test if writers get better at predicting their productivity by estimating the following regression, using measures of writers' beliefs about expected clicks per article prior to submission:

$$Accuracy_{it} = \alpha + \sum_t \beta_t Treat_i^j * \lambda_t + \beta_2 Accuracy_{i,0} + \lambda_t + \eta X_i + + \pi_s + \epsilon_{it} \tag{7}$$

where $Accuracy_{it} = |v_{it} - b_{it}|$ is writers i's average prediction accuracy in week t, $v_i t b_{i,t-1}$ is the average subjective belief about articles in week t stated when submitting the article, $\bar{v}_{i,t-1}$ is lagged average views per article to date $t-1$, and X_i is a vector of individual controls including tenure. We interact the treatment with week fixed effects to document post-treatment trends in prediction accuracy over time.

While writers in the treatment group may get better (or worse) at predicting the outcomes of their articles, learning the relationship between articles qualities and clicks requires experimentation. Thus, we expect a period following treatment in which writers engage in a process of "learning by doing" that may vary with writer experience and their risk tolerance. While the payoffs to writers in the fixed contract do not depend on adjusting any article characterstics, with the provision of performance feedback they may also adjust their writing on various dimensions due to intrinsic motivation or personal biases. To test the impacts of treatment on learning, we estimate the following regression by interacting treatment with week fixed effects:

$$y_{it} = \alpha + \sum_t \beta_t Treat_i^j * \lambda_t + \gamma Treat_g + \lambda_t + \eta X_i + \pi_s + \epsilon_{it} \tag{8}$$

where outcomes y_{it} include total views per week and average views per article per week. We expect $\beta_1 > 0, \beta_1 = \beta_t$ for all t if the writers know the relationship between article characteristics and views and respond to the pay-per-view contract immediately. However, if writers engage in a process of innovation, we expect lower treatment effects in early periods that diverge over time after a period of experimentation and feedback, $\beta_1 < \beta_2 < \beta_3$...

Selection

We will explore two main types of selection effects: selection into contract for writers randomized into Treatment Group B, and turnover of workers, which may vary between treatment groups.

The first selection effect will be estimated by regressinga_i, the probability of taking up PPV, on baseline characteristics, including risk aversion, pre-treatment average views, and $Accuracy_{i0} = |v_{i0} - b_{i0}|$, i.e. writers i's average prediction accuracy prior to treatment as in the following equation:

$$Pr(a_i|Treat_i = B) = \eta X_i + \epsilon_i \tag{9}$$

Generally, the average treatment effects from the introduction of pay-per-view contracts may be driven by changes in behavior (incentives) or selection (Lazear, 2000). We explore the role of selection through a change in the turnover of workers across the treatment groups (both quitting and hiring). To estimate the effects of treatment on the probability of quitting, we estimate the following Cox proportional hazards model:

$$log(h_{it}) = \beta Treat_i^j + \eta X_i + \pi_s + \lambda_t + \epsilon_{it} \tag{10}$$

where h_{it} is the quitting hazard of writer i at time t, and X_i contains baseline characteristics including average productivity to date, productivity beliefs, a measure of risk aversion, and ethnicity. The failure event (quitting) is defined as the final week in which writers submit an article following the introduction of the treatment.

We will also interact individual characteristics (C_i^j) with the treatment to test whether risk aversion, average productivity, overconfidence, and/or ethnicity predict quitting following introduction of the pay-per-view contract:

$$log(h_{it}) = \beta Treat_i^j * C_i^j + \eta X_i + \pi_s + \lambda_t + \epsilon_{it} \tag{11}$$

Further, following the introduction of the contract, new potential writers will be randomly assigned to a treatment group when they register as a writer. This random assignment to contract type may allow us to estimate whether assignment to pay-per-view contracts has differential effects for newcomers vs. those who had prior experience with the pay-per-article contract.

References

Bandiera, O., I. Barankay, and I. Rasul (2011). Field experiments with firms. *The Journal of Economic Perspectives 25*(3), 63–82.

Baron, D. P. (2006). Persistent media bias. *Journal of Public Economics 90*(1), 1–36.

Bénabou, R. and J. Tirole (2006). Incentives and prosocial behavior. *The American economic review 96*(5), 1652–1678.

Bolton, P., J. Scheinkman, and W. Xiong (2006). Executive compensation and short-termist behaviour in speculative markets. *The Review of Economic Studies 73*(3), 577–610.

Finan, F., B. A. Olken, and R. Pande (2015). The personnel economics of the state. Technical report, National Bureau of Economic Research.

Gentzkow, M. and J. M. Shapiro (2006). Media bias and reputation. *Journal of Political Economy 114*(2), 280–316.

Holmstrom, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization 7*, 24–52.

Lazear, E. P. (2000). Performance pay and productivity. *The American Economic Review 90*(5), 1346–1361.

Miller, G. and K. S. Babiarz (2013). Pay-for-performance incentives in low-and middle-income country health programs. Technical report, National Bureau of Economic Research.

Mullainathan, S., A. Shleifer, et al. (2005). The market for news. *American Economic Review 95*(4), 1031–1053.

Prendergast, C. (2002). The tenuous trade-off between risk and incentives. *Journal of political Economy 110*(5), 1071–1102.

Prendergast, C. (2015). The empirical content of pay-for-performance. *Journal of Law, Economics, and Organization 31*(2), 242–261.

Raith, M. (2008). Specific knowledge and performance measurement. *The RAND Journal of Economics 39*(4), 1059–1079.

Rantakari, H. et al. (2008). On the role of uncertainty in the risk-incentives tradeoff. *The BE Journal of Theoretical Economics 8*(1), 1–23.

Zabojnik, J. (1996). Pay-performance sensitivity and production uncertainty. *Economics Letters 53*(3), 291–296.