

Media and motivation: the effect of performance pay on writers and content*

Ivan Balbuzanov[†] Jared Gars[‡] Emilia Tjernström[‡]

January 31, 2019

Latest version [here](#). Version with tables and figures at the end of document [here](#).

Abstract

We study how incentives for journalists affect the quantity, quality, and composition of online media content. We report results from a field experiment within an online news firm in Kenya. Writers were randomly allocated to earn a piece-rate per article published or to a pay-per-view (PPV) contract. The PPV contract induced writers to produce more “popular” articles, but writers chose to submit fewer articles. Specifically, the PPV contract resulted in a 120% increase in total pageviews, a 180% increase in pageviews per article, and a 40% reduction in the number of articles produced. In line with our theoretical predictions, the effect on article quantity is concentrated among risk averse writers. Further, when given a choice, risk-averse writers tend to select out of the output-based contract. We also document changes along multiple non-incentivized dimensions of news production: writers shift away from producing local news towards national-level news. We see limited changes in article quality or in the prevalence of clickbait. Our study suggests that output-based incentive contracts have substantial implications for journalists’ effort and content choices, and more broadly for selection into risky “gig work.”

Keywords: Performance pay, Labor productivity, Media economics

*We thank Pascaline Dupas, Leah Lakdawala, Huailu Li, Teresa Molina, Laura Schechter, Alessandro Tarozzi, Maisy Wong and participants at the WGAPE 2017 annual meeting, Advances with Field Experiments Conference, the Annual Bank Conference on Africa 2018, as well as at seminars at Universitat Pompeu Fabra, Michigan State University, and UW-Madison for helpful comments and discussions. This study is registered in the AEA RCT Registry and the unique identifying number is: [AEARCTR-0002396](#). This work is supported by funding from the IFREE small grants program. Gars gratefully acknowledges financial support from the Policy Design and Evaluation Research in Developing Countries Initial Training Network (PODER, funded under the Marie Curie Actions of the EUs Seventh Framework Programme, Contract Number: 608109) The opinions expressed herein do not reflect the opinions of the sponsoring agencies. All remaining errors are our own.

[†]University of Melbourne, AUS; ivan.balbuzanov@unimelb.edu.au

[‡]University of Wisconsin-Madison, USA; gars@wisc.edu, tjernstroem@wisc.edu

1 Introduction

The news media plays an important role in the functioning of democratic economies, both by providing consumers with information about policies that affect individuals and firms, and by establishing citizen-government accountability relationships. These crucial roles of the media are especially important in developing countries, where weak institutions often leave large gaps in political accountability. While a growing literature investigates the effect of political and economic forces on media content at the newspaper level (see for example [Beattie et al. 2017](#) and [Di Tella and Franceschelli 2011](#)), little empirical evidence exists on the relationship between individual journalists’ incentives and the quality of the news articles that they produce.

The topic of journalists’ incentives is receiving increased attention ([Murtha, 2015](#)), perhaps due to the rising prevalence of performance contracts for online journalists (also known as “pay-per-click” contracts).¹ Supporters of the model argue that readers value investigative journalism and that giving writers greater discretion over how to connect with their readers should reward in-depth reporting. For example, Vogt et al. (2016) document that crowd-funded journalism projects have been increasing rapidly in recent years. In 2015, over 25,000 individuals contributed to such campaigns, and longer-form journalism stood out as the most commonly funded category, accounting for around 43% of crowd-funded projects.

Detractors counter by noting that journalists might instead choose to dumb down stories, increasing their use of easy topics and “click-bait” headlines. Despite growing concerns about the impacts of changing business models on the quality of journalism and on the polarization of political news, pay-per-click contracts have received limited attention from economists. To our knowledge, ours is the first study to rigorously explore how output-based performance contracts affect journalistic quality and (short-run) firm profits in the market for news.

Specifically, we employ a within-firm experimental design to study how pay-per-click contracts affect the quantity and quality of news content and whether different types of writers respond differently to performance pay. Our experiment took place within an online news firm that sources all its news stories from local, independent reporters. Writers therefore operate as

¹For example, media firms such as Slant and Politico have implemented bonuses for every 500 clicks on top of a fixed wage; the online publishing platform Medium now rewards writers based on the “engagement” that their articles receive.

independent “gig-workers” who can choose how much to work, what to write about, and how much effort to allocate to each story.

Quantifying the effects of journalists’ incentives on contracted and non-contracted tasks is important for several reasons. First, pageviews are an important source of advertising revenue for many online firms. Firms therefore care about the total number of clicks that their website receives and how much it costs to produce each click. Whether or not an incentive contract is able to produce more and cheaper pageviews is therefore a central question from a firm perspective. Second, online news production is a context where rewarding workers on the inputs into production is likely prohibitively costly. Output-based contracts are therefore more feasible, but may result in a reallocation of effort away from non-incentivized tasks. Third, the market for news is a context where distortions and bias can be particularly harmful—both for individual media firms and for the media’s ability to function as a conduit of information. If performance contracts shift the type, quality, or accuracy of news production, this could have long-term consequences for the public’s perceptions of media credibility.

In our context, workers choose how many articles to submit and how much effort to expend on each article on average. Our theoretical model predicts that writers will expend more effort under the PPV contract than under the piece-rate and that this increased effort will translate into higher expected pageviews. This comparison is harder to make for the optimal number of articles that writers choose to produce. Depending on a writer’s level of risk aversion and/or her marginal cost of effort, the optimal article quantity could be higher or lower under PPV compared with a piece-rate pay structure. The effect of PPV contracts on total production and on non-incentivized tasks are therefore empirical questions.

Our experimental design allows us to explore these questions. We randomly allocated writers to three contract groups: the first group (Control) received a piece-rate per published article; the second group was moved to a pay-per-view (PPV) contract with a piece-wise linear pay structure that rewarded writers based on the number of unique views that their articles received; writers in the third group were allowed to choose between the two contracts (Opt-in). We find that the PPV contract modestly increased per-article effort, sharply increased pageviews, and reduced the number of article submissions. In line with our theoretical predictions, the effects on article

quantity is concentrated among risk averse writers.

Specifically, we estimate intention-to-treat (ITT) effects, we find that the (PPV) contract decreased treatment writers’ article submissions by 7.3 articles per week relative to the control group (mean = 16.6), implying a decrease of over forty percent. Alongside the decrease in submissions, the output-based performance contract increased average views per article by 180 percent and increased treatment writers’ total views by 120 percent, relative to the flat rate. Self-reported effort per article increased by a modest minute per article from a baseline of roughly 37 (not statistically distinguishable from zero), and total weekly effort drops by 250 minutes per week (mean = 613) due to the decrease in submissions.

Figure 1 shows the distribution of average pageviews per article in the treatment and control groups pre- and post-treatment (with article views winsorized at the 99th percentile). Post-treatment, we can see that the distribution of article views has shifted quite dramatically to the right in the treatment group, and if anything to the left in the control group. Below we demonstrate that these treatment effects were immediate and quite persistent. Given how quickly the treatment changed the distribution of views, it appears that writers had knowledge about some aspects of the production function but were not sufficiently incentivized to allocate effort or to use the contextual information under the piece-rate contract.

In addition to the effect of the performance contract on article popularity and writer effort, we also care about the quality and content of the articles that writers produce—aspects that were not directly incentivized by the contract but that the principal cares about. The treatment had limited effects on article quality (as rated by the firm’s editors) or on the prevalence of “clickbait” titles. However, we find that writers respond to the PPV treatment by increasing the proportion of their articles that cover political topics by 29 percent (from a baseline of 35 percent) and that they contribute around 20 percent fewer articles on local (county-level) topics as a share of their submissions. This composition effect is likely driven by the fact that the PPV contract incentivizes writers to produce articles with the highest potential for views, leading to a decrease in local news coverage.

This shift away from local news suggests important potential drawbacks of pay-per-click contracts. Previous research shows that a decrease in the availability of local news leads to lower

voter turnout, fewer candidates running for office, and a lack of civic participation ([Oberholzer-Gee and Waldfogel, 2009](#); [Snyder Jr and Strömberg, 2010](#); [Gentzkow et al., 2011](#); [Schulhofer-Wohl and Garrido, 2013](#)). Absent direct incentives to produce a balanced portfolio of news, performance contracts may therefore have negative side effects. This result additionally highlights the difficulty of designing performance contracts for multi-dimensional tasks.

Our study contributes to the empirical literature on performance pay by producing evidence within a particularly important setting: the market for digital news. This market constitutes an ideal context for studying output-based performance contracts: productivity (i.e., views) is relatively easy to measure, inputs are not easily defined or observed, and writers may have incomplete knowledge of the marginal returns to different actions. Our contribution to the literature on incentive schemes in firms is threefold: First, our study is to our knowledge the first experimental evaluation of a switch from a piece rate to performance-based incentive contracts for a multi-dimensional task with stochastic output. Second, we are able to examine the effects of performance pay on firm profits due to the one-to-one relationship between article views and advertising revenues. Third, our experimental design allows us to directly examine the relative importance of incentives versus selection effects, two mechanisms that the literature has had limited success in separating.

Previous empirical studies of performance contracts within firms focus primarily on a switch from hourly wages to piece rates. By their nature, piece-rate contracts are common when tasks are simple or repetitive, inputs can be monitored (imperfectly), there is little scope for innovation or discretion, and output is not stochastic.² Productivity impacts from a switch to piece rates range between 20 and 50 percent, and the impacts on profits tend to be dampened by increased managerial costs. Given that we find substantially higher productivity increases moving from a piece rate to a performance contract, we interpret the difference as a result of journalists having discretion over their actions (delegation) and private information about the returns to actions that are insufficiently incentivized under the piece-rate contract ([Prendergast, 2002](#); [Raith, 2008](#)).

The evidence on performance pay also complements existing experimental research on how

²For example, [Lazear \(2000\)](#) studies workers in a firm that install windshields, [Shearer \(2004\)](#) studies tree planters in Canada, and [Bandiera et al. \(2005\)](#) study fruit-pickers in the UK.

to best incentivize public service delivery in developing countries. [Mohan et al. \(2017\)](#) compare input- and output-incentive contracts for healthcare providers' in Karnataka, India. They find that only healthcare providers with advanced qualifications reduced post-partum hemorrhage under the performance contract, presumably because they could apply their knowledge and implement context-specific strategies to prevent bleeding. [Muralidharan and Sundararaman \(2011\)](#) focus on performance bonuses for teachers based on student test scores in Andhra Pradesh, India and find that bonuses lead to significant increases in student performance (0.27 SD increase in math test scores). Using teacher surveys, they find that teachers innovated after the introduction of the performance bonuses by assigning more homework and giving practice tests, and allocating effort towards weaker students. In both cases, the benefits of performance pay are higher for multi-dimensional tasks when the production function is partially known by the principal, as is the case in the production of news.

We are also able to examine the effects of performance pay on short-run firm profits due to the one-to-one relationship between article views and advertising revenues. While performance contracts have been implemented in a variety of settings ([Lazear, 2000](#); [Finan et al., 2015](#)) to overcome principal-agent problems, the empirical evidence on whether it actually matters for overall firm profits is inconclusive ([Prendergast, 2015](#)).³ While agents face more risk in uncertain environments, performance contracts may be more efficient when agents have specific knowledge about the production function ([Raith, 2008](#)), or can learn the returns to specific actions. Further, it has been recognized that the marginal returns to delegation are higher when the principal cannot observe or define effort ([Prendergast, 2002](#)). We take advantage of the fact that firm profits in our setting are roughly proportional to pageviews, and estimate the impacts of contract structure on short-term firm revenue. The PPV contract group resulted in substantially higher views and revenues. In addition, lower submissions resulted in reductions in editor time costs. As a result of these impacts, the firm switched all writers to a PPV contract following the conclusion of the experiment.

However, in the longer-term, there may be negative consequences. Contracting on outcomes has been shown to crowd out intrinsic motivation ([Bénabou and Tirole, 2006](#)), incentivize short-

³See [Bandiera et al. \(2011\)](#) or [Miller and Babiarz \(2013\)](#) for recent overviews of empirical evidence on pay-for-performance contracting in developing countries.

termist behavior that rewards 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](#)). Alongside the increase in revenues, we have noted decreased local news production under the PPV contract. While we do not find immediate impacts on quality or click-bait, we are unable to rule out medium- and long-terms impacts that may affect the firm’s reputation and writer productivity.

Another key question in the literature has been whether performance contracts primarily impact workers through behavioral change or through selection effects ([Bandiera et al., 2011](#)). Empirical evidence separating these effects in personnel settings is scarce. [Lazear \(2000\)](#) observes an individual firm with a staggered roll-out of a piece-rate contract and finds that half of the 44 percent increase in productivity was due to more productive workers selecting in, i.e., joining the firm. Relatedly, [Bandiera et al. \(2007\)](#) find that productivity increases when managers’ pay becomes more closely tied to output, and attribute roughly half of these worker-level productivity increases to manager discrimination in hiring. However, as [Dohmen and Falk \(2011\)](#) point out, sorting may also take place along other dimensions than quality, potentially affecting the workforce composition in unanticipated ways. We therefore also examine the relative importance of incentives versus selection effects in driving the treatment effects that we observe. We study selection by comparing the characteristics of writers who opt in to the output-based contract to those who opted out. Risk averse writers are less likely to opt in to the performance pay contract. We also compare the characteristics of writers who remain active after the contractual change across control and PPV. Risk averse writers are more likely to reduce the number of articles that they submit than are risk tolerant writers; some writers reduce production to zero, effectively dropping out of the market.

In ongoing work, we plan to additionally examine whether performance contracts increase experimentation, i.e. the acquisition and productivity-increasing use of information about the marginal returns to writer effort. While some theoretical work focuses on contracting when the marginal productivity of effort is stochastic ([Zabojnik, 1996](#); [Rantakari, 2008](#)), empirical tests separating the information effect from the incentive effect of performance contracts remain limited.

2 Background

2.1 Media market

Kenya’s post-independence media landscape enjoys a fairly broad range of media outlets, with a mix of state-owned and private media outlets, most of which publish in English (Lohner et al., 2016; Freedom House, 2017). The news media consist of at least four daily newspapers, one business daily, and a handful of regional weekly newspapers (Freedom House, 2017). Although media coverage has traditionally been relatively rigorous, especially in the print sector, editorial pressure and the political preferences of advertising sources shape coverage at many outlets (Freedom House, 2017). Furthermore, reporters can face repercussions for critical coverage, including dismissal.

Additionally, human rights groups have expressed concerns in recent years that press freedom is in sharp decline (Namwaya, 2018; Freedom House, 2017).⁴ Some sources associate the reduction in press freedom directly with ownership concentration and media capture. Freedom House (2017) notes that most media outlets are owned by politically connected people or politicians, and that five media companies capture over 70 percent of all media consumers. The public’s trust in the media sector is, perhaps unsurprisingly, low. Simiyu (2014) surveys two small but broadly representative samples of citizens and reporters after the 2013 presidential elections, and finds that a stark majority of the Kenyan electorate believe the media was biased and partisan in their reporting of the election. The sampled journalists corroborated these perceptions, stating that their journalistic freedom was tightly limited during the coverage of the election.

2.2 The news firm

We study an online newspaper in Kenya that sources its stories from local reporters, who have been paid a piece-rate for each article they publish since the website’s launch in 2014. Writers need no previous qualifications to submit content, only a device and basic writing skills. The website averaged over 1.5 million active readers between May 1st and September 1st of 2017

⁴The most recent example occurred in early 2018, when the Kenya Communications Authority switched off a number of television stations for broadcasting live from the site of a political action by the opposition leader. The High Court of Kenya issued an order suspending the media ban, but the government stationed police around the relevant government offices to block court officers from serving the order. (Namwaya, 2018)

(the four months leading up to the experiment) and has experienced dramatic growth: average monthly readers grew from 68,000 in January 2016 to 887,000 in December that year. As can be seen in Figure 2, which depicts daily pageviews since 2014, views have declined since the end of the election period, but show a substantial overall increase since the beginning of 2016.⁵

After signing up, writers can submit articles for publication subject to a 500-word limit. Each writer has a “workbench” page where they can edit their profile, review and submit articles, and see their payment history. In addition to their workbench, each writer has a public profile that links to their published articles, listing the date of publication and the number of pageviews that the article has received. It is entirely up to writers what stories they want to cover, and how frequently they want to submit. Professional editors curate all incoming articles and choose which ones to publish, monitoring for defamation, plagiarism, hate speech, and a few more issues that guarantee rejection.

Articles are reviewed by editors in the order in which they are submitted. While the firm encourages timely submission of articles on breaking news events, articles must meet a minimum quality standard to be published. The primary reasons for rejection, in the order of frequency, are for being too similar to published articles, “not engaging enough,” plagiarized, and too informal. Editors have no discretion in the topics of articles that get submitted and neither the editors nor the firm suggest topics or content. Prior to the contract change, writers were paid 100 KES (about 1 USD) via the mobile money system M-Pesa if an article was published.⁶

In the month leading up to the contract change, 122 writers submitted articles and 107 successfully published at least one article. This resulted in an average of 830 published articles per week. In this month, the active writers were active for 4.3 weeks on average, where we define active as submitting at least one article in a given week. Conditional on being active, the average weekly number of submissions and publications were around 7 articles. However, there is substantial heterogeneity in the activity across writers in the numbers and rates of submission and publication. The median writer published only 2.35 articles per week while the top quartile

⁵The sharpest increase in pageviews coincides with the run-up to the Kenyan presidential election on August 8, 2017 (Election #1), where the incumbent President Uhuru Kenyatta won reelection with 54 % of the vote. The election results were challenged in the Supreme Court by the opponent, Raila Odinga, and the result was annulled under the condition that fresh elections would be held on October 17, 2017 (Election #2).

⁶Payment for published articles was disbursed weekly until June 6, 2017. Since then, writers have been paid every three days.

published almost 13 articles. Further, 40 percent of writers submitted in 3 or fewer weeks out of the two months, and only 30 percent submitted in all 8 weeks. Half of all articles published over the period were submitted by just 13 writers and one third of writers contributed five articles or less. This suggests overall that the “gig workers” that we study range from depending quite heavily on writing as a source of income, to very occasional writers. As described in more detail in Section 3, we stratify the randomization on various measures of participation to ensure balance across these different types of writers.

3 Experimental design

To examine the effects of an output based incentive contract for writers on the quality and quantity of their output, we randomly varied writers’ contractual terms for the treatment groups while maintaining the flat rate (status-quo) contract for the control group. Section 3.1 describes these treatments in more detail. Section 3.2 then describes the randomization, while Section 3.3 discusses how the performance pay contract was calibrated.

3.1 Experimental treatments

All writers who had published an article in the last 12 months since September 5, 2017 were randomly allocated into one of the following three treatment groups:

- *Control group/ Piece-rate*: The control group contracts maintain the status-quo contract of 100 KES (1 USD) per published article. This contract has been in place since the inception of the website in May of 2014. To receive payment in the control group, articles must pass the basic editorial bar (no repeat articles, no plagiarism, no hate speech, no defamation) and writers are paid within three days of publication.
- *Pay-per-view (PPV)*: The pay-per-view treatment group contract pays out based on a kinked and discontinuous piecewise-linear fee structure that is a function of the number of unique views that the article generates.⁷ The kinked fee structure is defined as follows:

⁷The exact calculation is based on the “users” metric in Google Analytics, which measures the number of *unique* views by IP address. This metric prevents writers from easily manipulating their payments by, for example, refreshing the page in their browser. We will use “users” and “views” interchangeably throughout the paper.

(i) there is a lower threshold of 400 views, below which writers receive no pay at all; (ii) between 400 and 800 views, writers receive a rate of 125 KES per 1000 views; and (iii) all pageviews beyond 800, writers receive a lower rate of 12 KES per 1000 views, with no upper limit. Section 3.3 provides more details on how the contract was calibrated. Payments are issued within three days of publication based on the number of users that have accrued at the time that payroll is processed. Writers can continue to be paid for an individual article in every following pay period if the article continues to attract users.⁸ This fee structure applies to all users after publication of the article and does not restart in each pay period.

- *Contract choice (Choice)*: The third group were allowed to choose between the status quo piece-rate and PPV. They were required to select their contract when they logged in to submit an article as soon as the intervention had launched. Writers in the choice group were informed that their selection would be permanent. The terms and payment timing of the two options are identical to those described above.

The firm informed writers of the new contracts via SMS on the day of the intervention launch. All writers received a message from the Head Editor on their workbench that updated them on the success and growth of the firm and thanked them for their participation in that success. This message was partly to ensure that any potential prompting effects of a message from the editor would be held equal across treatment and control, as well as to update all writers' current knowledge about firm growth. Writers in the treatment groups were told, in addition to the basic message, that they were to be paid according to a new contract, and that the change would take place immediately. The language to PPV writers included a message explaining that the firm was trying to find a better way to reward successful writers for their work. The writers were not shown the complete pay schedule, but rather a table noting the key kinks in the payout schedule (as seen in Figure 3).

PPV writers also received some information about the data that their payments would be based on (values from Google Analytics), about the very high correlation between “users” and the pageviews that they could already see on their profile pages, and about where they could

⁸Ninety-seven percent of pageviews are received within the first two days of publication on average.

locate their payment information for each article. Writers in the choice group received the same message, but were told that they could choose their contract by logging into their writer page and selecting their preference.

The compensation and responsibilities of the firm’s editors remained unchanged during the duration of the experiment and they were blind to the treatment status of writers. This precludes any confounding the effects of treatment with editor behavior. Editors are paid a fixed hourly wage for their editorial work but are also permitted to submit articles for publication. Editors that continued to submit articles were paid under the status-quo contract.

3.2 Randomization

All writers that had published at least one article in the year prior to the treatment (i.e., since September 5, 2016) were included in the randomization. In order to ensure balance on key baseline characteristics writers were stratified into 8 groups based on whether they had published at least one article in the month prior to treatment (0/1), if they submitted more than one article in the weeks when they were active (0/1), and if they were above or below median views per article for the period that they were active (0/1).⁹ Table 1 shows balance tests. Writers assigned to the Control, PPV and Choice treatments are well balanced along observables at baseline, including how long since they first signed up (in weeks), how many of those weeks they were active, the number of submitted and published articles, their average pageviews (both per-week and per-article), and their stated risk tolerance. Of the 480 writers included in the sample, twenty-five percent (172) published at least one article within two months prior to the intervention and forty percent (235) published within four months prior (not shown in the table).

Any new writers who signed up to write for the firm following the introduction of the treatment were also randomly assigned to a treatment group after registering as a writer. Advertisements on the website do not include information about the treatment or the pay structure and potential writers would only be aware of the available contracts if they knew one of the

⁹The views per article were time-demeaned to account for the growth in views over time; we worried that a non-demeaned views/article measure would introduce a correlation between views and the amount of time that had passed since the writer was last active.

current writers. We attempted to prevent writers from registering again to get a more preferred contract by requiring that no email address can be registered twice and that no phone number can receive payment for more than one email address.

3.3 Contract parameters and design

Figure 4 shows the actual payoff schedule that writers faced under PPV. The firm calibrated these contract parameters based on the previous months’ user data, with the dual aims of incentivizing high-impact articles, while ensuring that the *ex ante* expected payout would remain the same.¹⁰

The firm was also concerned with limiting the proportion of “losers” among current writers, which was the reason to introduce a steeper slope between 400 and 800 views. Of the articles used in the calibration, the firm chose values so that 44 percent of articles in the PPV contract would earn less than the 100 KES they would have earned in the status-quo contract (the firm’s definition of “losers”) and 27 percent of articles would have received no payment at all. Meanwhile, nearly 11 percent of articles would have received double the realized status quo earnings of 100 KES. Thus, writers in the PPV treatment face substantial downside risk conditional on an equal amount of effort, but the returns to increased effort have the potential to outweigh these losses. It is also clear from Figure 4 that the non-linearity introduced by the kinks suggest that risk neutral writers, assuming that they have close control over the number of pageviews that their articles receive, should target producing articles that get 800 views, since a single article needs to reach over 9,100 pageviews to receive the same level of payout as two 800-view articles. As we will see, writers seem to have some control over their pageviews but this control is quite imprecise and they have imperfect knowledge about exactly how many pageviews their articles will receive.

¹⁰Given the stochastic nature of output-based pay in the market for news, the literature suggests that risk averse writers may need to be compensated for the added risk. 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).

4 Theoretical framework

The model presented in this section is based on the classic linear-contract moral-hazard model with normal noise and CARA preferences (e.g. [Holmstrom and Milgrom, 1991, 1987](#)). One key difference in our setting is that we allow agents to choose the number of tasks to perform. It is very similar in the general question and approach to ([Butschek et al., 2017](#)). Our general model structure is very similar to theirs, but their model introduces more parameters in the cost function, and their focus is on a change in the per-item pay within a high-powered contract. Further, our results differ substantially from theirs.

Assume that the decision maker chooses how many articles to write and submit ($n \in \mathbb{N}$) and how much effort to expend on each article on average ($e \geq \underline{e}$). We interpret \underline{e} as the lowest level of effort that guarantees meeting the editorial standards and subsequent publication. In what follows, we normalize $\underline{e} = 0$. The agent has utility with constant absolute risk aversion over her monetary payment minus the cost of effort. The cost-of-effort function $c : \mathbb{R}_{++} \times \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$ is $c_1 ne^2 + c_2 n^2$ for strictly positive parameters $c_1, c_2 > 0$.

Conditional on a chosen effort level e , the effective impact of each article is a $e + \eta$, where η is a normally distributed variable with mean $m > 0$ and variance σ^2 , and agents know both m and σ . Thus the expected impact of each article equals $e + m$. The article’s impact is proportional to the number of pageviews it attracts.¹¹ This implies that for a chosen production plan (n, e) , the total impact of all articles is $ne + \varepsilon$, where ε is a normally distributed variable with mean nm and variance $n\sigma^2$. To avoid uninteresting discreteness issues, in the next version of our model we will assume that the number of articles can be any non-negative number: i.e. $n \geq 0$.

The payment contracts we consider can be modeled by a pair of parameters (α, β) , where α is the fee paid per approved article and β is the scaling parameter that determines the quality-contingent payment. If α is positive, we can normalize it to equal 1. Thus, if the quality of an article is q , the payment for the agent for that article is $1 + \beta q$.

We separately consider the agent’s optimization problem when faced with a flat contract (where $\alpha_{FC} = 1$ but $\beta_{FC} = 0$) and a linear (fee-per-pageview) contract (where $\alpha_{LC} = 0$ but

¹¹For the time being, we are assuming that the marginal product of effort is fixed and known, and the only way for agents to increase the expected number of pageviews is to increase their effort. We will add an experimentation aspect to the framework, where certain possible action have ex-ante unknown marginal productivity of effort.

$\beta_{LC} > 0$). This allows us to state our first result, with ρ denoting the coefficient of absolute risk aversion.¹²

Proposition 1. *The utility-maximizing production plan under the flat contract $(1, 0)$ is $(n_{FC}^*, e_{FC}^*) = (\frac{1}{2c_2}, 0)$. The utility-maximizing production plan under the linear contract $(0, \beta)$ is*

$$(n_{LC}^*, e_{LC}^*) = \left(\max \left\{ 0, \frac{1}{2c_2} \left(\frac{\beta^2}{4c_1} + \beta m - \frac{\rho}{2} \beta^2 \sigma^2 \right) \right\}, \frac{\beta}{2c_1} \right).$$

Proposition 1 implies that writers will choose more effort under the linear than under the flat contract. This comparison is harder to make for the optimal number of articles under the two contracts. If risk aversion and/or the marginal cost of effort are relatively high, it is optimal for the agent to choose small n_{LC}^* or even $n_{LC}^* = 0$ under the linear contract. As discussed below, depending on the parameters of the model, it is possible that the agent also chooses to produce *more* articles under the linear than under the flat contract.

In what follows, we assume the following relationship between the parameters of the two contracts:¹³

$$1 = \alpha_{FC} = \beta_{LC} m, \tag{1}$$

where we use the fact that the expected quality of each article under $\underline{e} = 0$ is equal to m . Before stating the next result, note that with this parameterization if, say, $\rho = 0$, the agent would choose to write strictly more articles under the linear than under the flat contract.

Proposition 2. *There exists $\bar{\rho} > 0$, such that for all $\rho < \bar{\rho}$, the agent prefers the linear contract $(0, 1/m)$ over the flat contract $(1, 0)$. Conversely, for all $\rho > \bar{\rho}$, the agent prefers the flat contract over the linear contract.*

Several testable implications emerge from the model: first, we expect average per-article effort to increase under the linear contract. This should translate into higher average pageviews for

¹²Proofs of Proposition 1 and 2 are presented in Appendix A.

¹³This is based on the way the per-pageview payout for the linear contract (i.e. β_{LC}) was determined by our partner firm. The firm chose β_{LC} so that the total expected per-article payout under the linear contract if calculated using the average per-article pageviews under the flat contract would equal the flat-contract per-article payment $\alpha_{FC} = 1$.

treatment writers. Second, the number of submissions under the PPV contract should diminish in writer risk aversion, but risk aversion should not affect article quantity in the control group. Third, writers should sort based on risk aversion, with only sufficiently risk tolerant writers choosing the performance contract.

5 Data

This section describes the data that we use to analyze our various research questions. We draw from several data sources: the main source is the firm’s administrative data (described in Section 5.1), but we also have access to the responses to a battery of questions that writers must complete when they submit an article, and another set of questions that editors must complete when they review an article (described in Section 5.2). We further generate additional data using readability score algorithms and crowd-sourced ratings and evaluations of the articles (still to be done, but an outline can be found in Section 5.3). We then describe our primary outcome variables in Section 5.4.

5.1 Observed measures

The digital submission framework and website data allow us to match page-level measures from Google Analytics (GA) with individual article-level and writer-level information from the firm’s database. The firm’s database and Google Analytics data begin in early 2014. For the purpose of this study, all writer and output data are collected beginning in September of 2016, with the exception of writer tenure at the firm. As discussed in Section 3.2, all 480 writers that published at least one article since September 5, 2016 were included in the randomization. Our dataset represents all articles and output data for the firm except for articles published by editors.¹⁴

GA provides the data for one of our key outcome variables: article pageviews. The measure that we call pageviews is labeled by GA as “users;” this measure constituted the basis of the performance pay contract.¹⁵ We also obtain several secondary outcome variables from GA: first,

¹⁴Editors that also publish articles were excluded from the study because they submit primarily live news feeds (such as election updates) and have a privileged position in the publication process. On average, editor articles perform higher than the writer average but are published much less frequently.

¹⁵“Users” measures the number of people (IP addresses) that have viewed a page at least once in a given day,

we observe where an article’s pageviews are coming from (“source” in GA) which includes the number of times an article is accessed from outside sources including Facebook and Twitter. From this we construct the share of an article’s views that come from social media. Second, GA reports the article-level bounce rate (a bounce is when a person leaves the website from the landing page without browsing further). Third, GA allows us to examine the amount of time that a reader spends reading an article, which is one potential measure of reader engagement with the article.

The firm indexes log files for writer profiles, articles, and payments using ElasticSearch. We use this data to create writer- and article-level measures of input decisions (frequency of submission, etc.) and article characteristics. Specifically, we obtain each article’s date and time of submission and publication, its title, text, revision history, category (politics, news, etc), and location.¹⁶ Elastic Search also contains information on the writer’s tenure with the firm, and we construct other writer-level variables, such as average weeks active. We subsequently match this data to the GA article-level data using page titles.

5.2 Elicited measures

In addition to the administrative data, we elicit several measures from the writers each time they submit an article, and from the editors each time they process an article. Writers are prompted upon submission to (i) approximate the number of minutes that they spent putting together the news story and (ii) predict how many pageviews they think the article will get in its first week. Writers are assured in writing that editors cannot view their answers to these questions and that the responses will not affect an article’s publication chances. Rather, they are encouraged to report truthfully so that “the firm can learn to better serve them.”

Each writer has a profile page that provides pageview counts for each article that they have published, which they and any readers can access by clicking on the profile link in an article’s byline. Given that writers can access article-level pageviews and that users and pageviews are

while pageviews measures the number of views on a page by session. For a single page, these numbers differ whenever a single user has viewed the same page more than once and only if that is done in more than one session (sessions restart after thirty minutes of inactivity). The two measures are highly correlated ($\rho = .987$), but users is more difficult to manipulate.

¹⁶The firm allows writers to tag articles both with a topic and with a county

highly correlated, we argue that writers’ predicted pageviews constitute a reasonable measure of writers’ expectations about the profitability of an article.¹⁷ Collection of beliefs and effort data began on July 27, 2017. We therefore have a little over a month of data on expectations prior to the introduction of the treatment (for the 127 writers that submitted during this period).

We also asked editors a set of questions about articles that they reviewed prior to publication or rejection. First, they were asked to rate the article on a scale from 1-5, with 5 representing the highest quality and 1 the lowest. Second, editors also had to predict how many pageviews they thought the reviewed article would get in its first week. We believe that editors took this task reasonably seriously, and will use the first measure as our first measure of article quality.

Finally, we elicited writers’ “general willingness to take risks” via an optional online survey. While this measure is not directly related to a theoretical risk aversion parameter, it has been shown to be a good predictor of actual risk-taking behavior (Dohmen et al., 2005). Note that the optional nature of the survey resulted in a relatively small sample size. Fifty to sixty percent of the writers in each treatment group completed the survey. Survey completion is not significantly different across groups, but writers who completed the survey were more active producers on average. Our risk aversion results should be interpreted with this in mind.

5.3 Generated data

5.3.1 Readability scores

To explore the effects of the contract change on content, we construct several measures: first, we apply readability formulas to the article text to construct different “readability” measures that allow us to examine if writers change something about how they write following the introduction of the PPV contract. “Readability” generally refers to the level of vocabulary and sentence structure (Chall and Dale, 1995; DuBay, 2004), and these measures employ various evaluation approaches. Most measures use some function of word frequency and sentence length to construct a text-level score. To measure readability, we focus on the number of syllables, the number of

¹⁷We did not include financial incentives to encourage accurate time and belief responses. Effort reporting could not be incentivized, as we would be unable to verify reports. The evidence on how much incentives improve the accuracy of reported beliefs is mixed. In a recent study, Hoffman and Burks (2017) compare predicted productivity for truck drivers (miles per week) using both incentivized and unincentivized elicitation. They found no evidence that reported beliefs were different when drivers are rewarded for accuracy.

words, and three common and reliable scores in the literature: Flesch Reading Ease, Coleman-Liau, and Dale-Chall.

The Flesch Reading Ease is based on syllable counts, the Coleman-Liau index uses character counts (letters) relative to the number of sentences, and the Dale-Chall formula compares the text to a list of 3,000 “easy” words. The Flesch Reading Ease formula is measured on a scale from 0 to 100, with zero being the most difficult to read. The other two measures use textual characteristics to generate grade levels based on an estimate of the number of school years needed to comprehend the text. We take the negative of the latter two values so that higher numbers indicate writing that is easier to read across all scores. While the formulas used to calculate the scores vary in their constants and in the weight assigned to different formula components, they generally provide similar rankings across texts.¹⁸

5.3.2 Clickbait scores

Clickbait scores are generated using a deep learning algorithm that classifies headlines using a training dataset that includes 12,000 headlines that are allocated equally across category (clickbait vs non-clickbait). The headlines are classified based on their source. Clickbait headlines were taken from BuzzFeed, The Huffington Post, etc. and non-clickbait headlines were taken from sources including The Economist, The Wall Street Journal, The Guardian, etc. The algorithm is imperfect for the setting (trained on US outlets like BuzzFeed and the New York Times), and we are in the process of re-training it on Kenyan equivalents.

5.3.3 Bias and quality ratings

Concerns about the implications of performance pay in the media often center on the threat of partisanship and its potential ramifications for polarization and political accountability. We are therefore keen to explore the effect that these new contracts have on journalistic slant and bias. Bias and slant have been used interchangeably in the literature to refer to the degree to which a news outlet or journalist presents information in favor of a political party or ideology ([Gentzkow and Shapiro, 2006](#)). In this paper, we will identify the difference in overall ideological slant

¹⁸See [Hengel \(2016\)](#) for further discussion on the comparability of the scores and the criticisms of their accuracy by context.

across contract types and focus on two specific types of bias. Bias is generally delineated into two types: issue framing which is the ideological slant of individual issues, and issue filtering which is the selective coverage of topics. Analyzing this effect of the intervention will also allow us to explore the relative strength of demand-side vs. supply-side sources of media slant and bias. 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 world view (Baron, 2006), while demand-side theories argue that bias only occurs in equilibrium if firms find it profitable to distort information to match consumers’ ideologies (Gentzkow and Shapiro, 2006; Mullainathan and Shleifer, 2005). Since the control group has little incentive to respond to reader preferences, we assume that any increase in bias or polarization in the treatment group would be due to demand-side factors, i.e. the political preferences of consumers.

To measure slant and bias, we use a crowd-sourcing technique similar to that described in Budak et al. (2016). We will enlist auditors on MTurk to classify the articles into politics or not-politics. Auditors will then be asked to identify 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?” using a 5-point scale. Answers will then be averaged across auditors (three per article) and normalized to generate party-favorability scores for each article in the politics category. We will use the MTurk human annotations as the labeled training set. Specifically, we plan to represent each article as a feature vector using state-of-the-art word embedding. The best classifier will be selected, and deployed to classify a larger set of news articles.

We will also use these annotations to classify writers. Writers choose the topics that they write about, so a writer who chooses to write about topics that are consistently negative towards a specific party may be intrinsically biased. However, negative news or events are time-specific, allowing us to categorize writers based on whether their articles are consistently more negative about a party than the average article that week. 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 that party than the average article that week. 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 whether

their coverage is perceived as fair or slanted. We will then separate articles into topics and compare the ideological slant between treatment writers and control writers to determine the degree of issue framing. Finally, by separating articles into topics, we can measure the number of articles written on each topic and compare the distribution of topic coverage across contract type to determine the degree of issue filtering.

5.4 Outcomes

We estimate treatment effects using writer-weeks as the primary level of observation.¹⁹ Writer-level measures are constructed using writer values at the week level. For example, the number of published articles for writer i at time t_1 is calculated as the count of all published articles for writer i in the first week post-treatment ($t = 1$).

We begin by evaluating the implications of the new contract for the firm’s objective function by exploring the average pageviews that a writer’s articles receive. We additionally compute the cost per click as a measure of the profitability of a writer’s output. We further examine the total number of views that treatment writers obtain compared to control writers. We further examine writer behavior by looking at writer effort expended (both per-article and per-month), as well as the number of articles submitted and published. The effects on effort per article, the number of articles submitted, and the interaction with risk tolerance relate directly to the testable implications of Proposition 2.

For the non-incentivized dimensions, we examine several aspects of article type and content. We examine editor-rated quality (described in Section 5.2), a click-bait score, the share of a writer’s articles that are political in focus, and the share of articles that they write on local topics (as opposed to national-level topics). We also examine five different readability scores (described in more detail in section 5.3). Finally, we examine whether readers interact differently with treatment writers’ articles. We look at how readers arrive on the article page (from social networks or not), the bounce rate, as well as how long readers spend on the page.

¹⁹This week-level regression is what we pre-specified. Readers could be concerned about the changing sample composition between weeks, as writers drop in and out of being active between weeks. We therefore additionally present our main results at the month level in Appendix B.

6 Results

6.1 Graphical results

We begin by showing graphically some of the key outcome variables by treatment group during the weeks prior to and following the treatment. Figures 5 - 10 indicate the beginning of the treatment period with a vertical dashed line. The triangles show weekly averages in the treatment group, and the hollow circles show averages in the control group. The dashed and solid lines denote local polynomial fits for treatment and control, respectively, with a 95-percent confidence interval. Figure 5 shows the average per-article pageviews. This figure clearly shows that the PPV incentives resulted in high-impact articles on average, with immediate and persistent effects.

The result in Figure 5 is in line with our theoretical predictions and we expect these high-impact articles to result from a higher per-article effort. Figure 6 shows that this is not the case.²⁰ How, then, are writers achieving these higher per-article pageviews? Perhaps they are writing higher-quality articles (or more scandalous, splashy articles), or catchier titles? We will return to this question below and in Section ??.

Figure 7 compares the total article views in the control and treatment groups before and after the treatment. It is clear that the PPV contract resulted in an immediate, sharp jump in the total number of pageviews for the treated writers, that persisted at least for the first half of the treatment period. The impacts on total pageviews appear to diminish towards the second half of the period.

Total pageviews are of course also a function of the number of articles published. Figure 8 looks at the average number of articles submitted. The treatment writers submit drastically fewer articles on average, suggesting that the writers in our sample are relatively risk averse.

We can also examine some of the questions about the impact of performance pay on content graphically. Figure 9 depicts the average editor ratings for the two contracts. The graphical results suggests that editors generally seem to evaluate the PPV articles to be of higher quality.

²⁰Figure 6 and 9 have a shorter pre-treatment duration than the other graphs as we did not elicit writers' effort levels or editor-rated quality until four weeks before the treatment. We also omit the confidence intervals from the pre-treatment period in Figure 6 as they are uninformative.

The graphical results suggest that concerns about performance pay leading to lower-quality news may not be justified in this context—if anything we observe the opposite effect. Writers’ payments are determined by pageviews rather than by some proxy for quality, but it is possible that treatment writers produce higher-quality articles (as rated by the editors) in order to reduce their likelihood of rejection.

We can also examine the clickbait score that articles receive, as measured by an algorithm that is admittedly imperfect for the setting. Figure 10 shows these results. Strikingly, the two groups appear very similar on this metric, but we don’t know whether this is measurement error due to an improperly trained algorithm or because the two groups are truly similar along this dimension. Note, however, that in ANCOVA regressions, the writer’s baseline clickbait score is a statistically significant predictor of current clickbait, suggesting that the measure is not pure noise.

The next section tests the theoretical hypotheses from Section 4 more rigorously, controlling for time and strata fixed effects. Section 6.3 explores whether these impacts are driven by behavioral changes by writers or by writer selection.

6.2 Average treatment effects

We begin our analysis by estimating the average impact of assignment to the PPV contract on writer submissions and article impact. Relative to a piece-rate contract, we expect contracts that reward outputs to be more efficient if agents have specific knowledge of the production function and if 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 PPV contract changes writers’ behavior along a number of dimensions and report both the intent-to-treat (ITT) estimates (i.e. simply comparing the treatment effects across groups) and the effects on active writers.

Following McKenzie (2012), we pool the data and estimate ANCOVA models to estimate the average treatment effects, controlling flexibly for baseline outcomes $\bar{Y}_{i,pre}$, individual characteristics X_i , fixed effects ν_s for the eight strata, and week fixed effects δ_t for $t = 1, 2, \dots, T$

post-treatment weeks.

$$Y_{it} = \sum_t \delta_t + \gamma_1 PPV_{it} + \phi \sum_{k=-4}^{k=-1} \bar{Y}_{i,k} + X_i + \nu_s + \epsilon_{it} \quad (2)$$

where Y_{it} is computed at the writer-week level. When we estimate impacts on total submissions and total views, we set values to zero in weeks during which a writer did not submit anything (i.e., the sample is balanced). When we look at per-article measures, we drop writer-weeks in which writers did not publish (or submit) articles. The pre-treatment outcomes ($\bar{Y}_{i,pre}$) are writer-level averages in each of the four pre-treatment weeks. Appendix B reports results estimated at the month-level, where $\bar{Y}_{i,pre}$ is simply an average over the last pre-treatment month. We cluster our standard errors at the writer-level (the level of randomization) in all estimations.

6.2.1 Average treatment effects on pageviews, effort, and submissions

Table 2 reports the ITT estimates for the full sample of writers over the post-treatment weeks. All outcomes are reported at the writer-week level. Per-article views are 180% greater in the PPV group than in the control group (column 1), and the cost per click is dramatically lower (column 2). This suggests that the per-article productivity is much greater in the treatment group. Total views in the PPV treatment group (column 3) are 120% greater than those in the control group. Note that this is despite the fact that we assign zeros to those who publish nothing in a given week. When we restrict the sample to weeks in which writers publish (column 4), the impact is of a similar magnitude in percentage terms.

Our model implies that this increase in pageviews would come from an increased per-article effort under the PPV contract, conditional on the production and cost functions specified (Proposition 2). In Table 3 we test whether total effort and effort per article indeed increase in response to the PPV contract using the self-reported measure of the number of minutes spent producing each article. While per-article effort appears to be unaffected (columns 1 and 2), we do see a sharp drop in weekly effort in columns 3 and 4.²¹ As with total pageviews, this is driven by the

²¹The effort regressions in Table 3 present results where self-reported minutes have been winsorized at the 1st and 99th percentiles to reduce the influence of observations with inordinately high or low minutes due to the

reduction in article quantity.

Turning to Table 4, we can see that submission behavior corresponds well with the graphical analysis in the previous section. Writers in the PPV treatment submit around 7 fewer articles per week (columns 1 & 2), and published roughly 6 fewer articles than did control writers (compared to a mean of around 17 and 14, respectively). The ITT estimates that include weeks in which writers did not submit at all (columns 1 and 3) are of similar magnitude as the estimates in columns 2 and 4, which suggests the effect of the treatment on submissions is not primarily coming from adjustments on the extensive margin at the week level. As our theory also predicts differences in submission behavior across risk aversion, the next section examines whether the PPV impact on submissions varies by our measure of risk tolerance.

A first pass at heterogeneous impacts can already be gleaned from Tables 2 and 4): it appears that more experienced writers (as measured by a writer’s tenure with the company, i.e., the number of weeks since they first joined) have higher total views but the same amount of per-article views. This seems to be driven by the fact that longer-tenure writers submit more articles. It is of course not clear whether early-joiners are simply different from more recent writers, or if this is actually a causal effect of experience.

The results from Tables 2, 3 and 4 combine to provide evidence that the treatment contract increased per-article impact, as predicted by the model. The substantial difference in cost-per-click suggests that the PPV contract had a positive impact on firm net revenues. We also observe a drop in the number of submissions in the PPV group and no clear impact on per-article effort. It is worth noting that these results are partial-equilibrium effects and only provide suggestive evidence of the true impact of the treatment on firm revenues and profits. For example, we are unable to clearly determine how much of the increase in productivity in the PPV treatment was due to substitution of readers away from articles in the control group or due to attracting readers from other sources. In Section 6.3, we will show that readers of PPV articles are more likely to arrive to the site from social media, which we take as suggestive evidence that the increased readership is not entirely driven by substitution.

self-reported nature of the data. The affected outliers report either 1 or over 10,000 hours per article. The results are robust to including these outliers.

6.2.2 Heterogeneous treatment effects on article submissions

Table 5 reports results that interact the treatment dummy with our measure of risk tolerance. Note that our theory predicts that the number of articles that writers produce should be decreasing in risk aversion; here this means that article quantity should increase in risk tolerance. Columns (1) and (3) report linear interactions, while columns (2) and (4) allow for the potential of non-linear effects with a quadratic interaction.

The theoretical predictions seem strongly borne out by this: more risk averse agents submit fewer articles on average, but this seems mostly driven by risk averse agents in the treatment group. The coefficient on the interaction between treatment and risk tolerance is both statistically and economically significant. It does not appear that the effect of risk tolerance on treatment writers’ optimal quantity is nonlinear, as the quadratic terms are small and insignificant. Figure 11 substantiates this by plotting the average marginal effects by risk tolerance: the effects on submissions are concentrated among risk averse treatment group writers and look close to linear.

Taken together, these results suggest that performance pay will have differential impacts on different types of workers. In a context like ours, where agents get to choose the number of tasks in addition to the per-task effort, firms may need to take risk aversion into consideration as writers may choose to drop out if faced with more risk than they are willing to bear. This could both have implications for the types of workers that remain—of great importance in news production—and for total revenues, if otherwise productive writers drop out entirely by choosing to produce zero articles.

6.3 Mechanisms

We now turn our attention to secondary outcome variables that can help shed light on the ways in which treatment writers are achieving higher impact. While writers were not directly incentivized to change the content of their articles, a change in the type of articles produced could explain how the workers in our sample were able to increase pageviews without increasing their effort. Perhaps they could produce “popular” articles by simply changing focus, or by adding catchy titles to their articles.

6.3.1 Content

Table 6 reports results on editor-rated quality, clickbait, and topic choice. Unlike the graphical analysis, which suggested that treatment writers received slightly higher ratings, column (1) provides little evidence that the PPV contract affected editor-rated article quality. While the point estimate on the treatment is positive, it is small and not significantly different from zero.

Similarly, the impacts on click-bait are statistically indistinguishable from zero (column 2). This could certainly be due to measurement error, as our algorithm is not perfectly trained for this context. That said, if some of the more pessimistic predictions by some media analysts were true—that quality will automatically plummet when journalists get rewarded based on views—we would expect to be able to pick something up. Further, the significant coefficient on baseline clickbait score suggests that there is some persistence in this measure over time, and that it measures more than pure noise.

Column 3 of Table 6 shows a sharp uptick in the share of political articles that treatment writers produce, which may be related to the timing of the treatment period around an intense political time in Kenya that increased public interest in political news. In another context, the category choice could be different (say, entertainment). Because their payment is contingent on views, treatment writers likely shifted their efforts towards political articles as a response to the high demand. Political articles make up a large proportion of all submitted articles for this period, and they also tend to receive higher average views than all other categories both before and after the treatment. Finally, column 4 shows that treatment writers submit fewer county-level articles. This is not unexpected, as the national audience is larger.

This shift in topical focus might reflect that PPV writers either have or are gaining knowledge about the average returns to political articles relative to other categories. We include week fixed effects to reduce the influence of the increased political activity during the time surrounding the election period. Conditional on a fixed number of political stories in a given week, PPV writers may try to submit high-impact political stories faster than control group writers. Writers in the piece-rate contract do not know about the existence of the PPV treatment group, so it seems unlikely that this effect reflects a substitution away from political articles in the control group. However, it could be that increased competition to submit timely political articles could lead

control writers to decrease the share of political articles they submit.

6.3.2 Readability and reader behavior

As we saw in Section 6.2, the PPV contract induced treatment writers to produce higher-impact articles. Given that the effects were immediate and persistent (especially for per-article pageviews), it seems as though writers already possessed knowledge about the types of inputs or actions that could increase pageviews. Here, we want to try to understand what these inputs or actions might be, beyond the shift towards political and national-level topics. Do PPV writers write more clearly? Do they share them via social networks post-publication? And do readers interact differently with PPV writers’ articles?

Table 7 addresses the first question by estimating treatment effects on article readability. The scores here are scaled so that articles with higher scores are easier to read. The first two scores, which are based on simple counts of the number of syllables and the number of words in the article, are not affected by the treatment. Furthermore, the overall regression fit for these measures is quite low.

For the three other scores, we can see that the treatment writers employ more difficult—or perhaps more sophisticated—language in their articles. This could be partly due to the greater share of politics articles, as we may expect political articles to use more “difficult” words than sports or entertainment articles. More experienced writers appear to write in simpler language, at least when we look at the Dale-Chall or Coleman scores. It also appears that writers are somewhat consistent in their writing style, as their baseline readability scores are significantly correlated with post-treatment readability scores. On its face, this would appear to be further evidence against the notion that performance pay will automatically lead journalists to produce “dumbed-down,” simplistic writing.

Table 8 examines whether readers interact differently with treatment writers’ articles. While the writer effort results in Section 6.2 suggested that treatment writers reduced the number of minutes spent preparing their articles, this self-reported measure does not pick up efforts devoted to post-publication sharing or publicizing of the work. The results in column (1) show that the proportion of pageviews that land directly on the article page from a social network (when the

reader arrives on the article through Twitter or Facebook) is significantly larger (by around 50 percent) in the treatment group. This could be indicative of post-publication effort by PPV writers (i.e., sharing their articles on social media), which could help explain the puzzle of no impact on writer effort. This is not the only explanation, of course: an alternative explanation is that readers are more likely to share PPV articles with their networks because they are more appealing for whatever reason.

The results in column (3) examines the bounce rate of article pages. A reader bounces if they leave the parent website after viewing only the page they landed on. We can see that the PPV articles have a lower bounce rate than the control group articles, but the point estimate is noisy. Bounce rates are often used as a measure of reader engagement, with lower bounce rates interpreted as being positive, as readers continue perusing the relevant site. We do not see any effect on the time that readers spend on a given article’s page—another measure of engagement.

6.3.3 Selection

The theory in Section 4 argues that the PPV contract will induce risk averse workers to submit fewer articles and to change their average effort for the articles that they do write. It also predicts that agents who are sufficiently risk averse could choose an optimal article quantity of zero under the PPV contract, i.e. the contract could introduce selection. If given the option, it also suggests that sufficiently risk averse writers should prefer the piece-rate contract over the performance-based contract. Understanding selection could matter if risk averse agents also write different articles, cover different topics, or are more or less susceptible to potential demand-driven bias. We also want to analyze selection to understand whether our treatment effects are driven primarily by the incentive effect or whether selection is a major factor.

Output-based pay might also be particularly likely to deter low-productivity writers, those with a high cost of effort, or those with a distaste for pay variation. On the other hand, the new contract could also induce writers who did not plan to submit again under the per-article contract to continue writing. Of these various explanations, we can only directly test a few. We already showed above that reductions in article quantity were primarily driven by risk-averse writers, suggesting that drop-out might be similarly concentrated among risk averse writers.

But what about when we give writers a choice between contracts?

Table 9 speaks to this question. The first three columns present marginal effects from a probit regression; the last three are from a linear probability model. We can see that our measure of risk tolerance is strongly predictive of a writer's of choosing the PPV contract across the different models. Note that only 14 percent of writers opted-in, and that the sample size is small, but we take this as strongly suggestive that risk tolerance matters. A one-unit increase in the risk tolerance measure (which was elicited on an 11-point scale) is associated with an increased probability of opting in of between 0.35 and 0.43. From OLS it seems like a one-unit change in risk tolerance would be associated with an increased probability of 7.7 percent, which is about half the actual opt-in rate.

We can also observe that writers with more experience seem to be less likely to opt-in. We do not know whether that is due to inertia and reference dependence, or perhaps due to better information about the riskiness of the PPV contract. Perhaps surprisingly, the average pageviews in the pre-treatment period do not seem to explain opt-in behavior; it seems like risk aversion is explaining much of this behavior. Again, these results should be interpreted with caution since relatively few writers responded to the survey and only a small proportion of writers opted in to the PPV contract.

Our second look at selection follows Lazear (2000), who includes worker fixed effects in a regression of daily productivity before and after a contract regime change. The Lazear study finds that including fixed effects reduces the estimate of the contract change by more than half; the author argues that the resulting estimate is the incentive effect. In the same vein, Table 10 includes writer fixed effects in a difference-in-differences (DiD) regression of views per article, cost-per-click, and effort. Columns (2), (4), and (6) include writer fixed effects. The effect of the PPV contract on per-article views is actually larger than when we do not control for the writer FE. The point estimates on cost per click and submission quantity are negative and statistically significant, and of very similar magnitude as the ANCOVA results. Overall, these results suggest that our results are largely due to incentives, as opposed to selection.

7 Conclusions

Using random variation in assignment to contract type for journalists in an online news firm in Kenya, we show that an output-based incentive contract sharply increased average and per-article impacts, and substantially reduced the frequency of article submission. We show that risk averse writers reduce their production much more steeply, suggesting that performance pay has heterogeneous impacts on different types of workers. This has important implications for online labor markets and other “gig worker” contexts, as the incentives may drive otherwise productive risk averse workers out of the market.

While we find that the performance contract increased both total and average article-impact, we are not yet able to determine whether the change will lead to an increase in long-term profits, although the much higher cost-per-click in the control group suggests that PPV is advantageous for the firm. In fact, the firm has since the experiment moved all writers to a revised version of PPV.

The new contract clearly led to an increase in article popularity, but a key remaining unknown in our analysis is the relationship between total content and total views. We cannot directly address what would happen if the number of articles produced drops “too far.” Furthermore, the increase in views per article in the performance treatment could be partly the result of

substitution of views away from the control group (competition). In future work, we plan to determine which effect dominates in the short- and long-run.

Ongoing work will pursue a deeper examination of the political content that constitutes the majority of articles produced during the treatment period. We hope that this will boost our discussion of how performance pay may contribute to (or reduce) bias and polarization in the market for news. In addition to trying to retrain our machine learning algorithm on Kenyan news sources, we plan to recruit workers on Amazon’s Mechanical Turk to evaluate title and text discordance and political bias for a subset of articles. Using this training set, we will use machine learning approaches to generate measures of click-bait and political bias for the whole corpus of articles. We will test the degree to which potential bias is driven by demand-side or supply-side bias.

References

- Bandiera, O., I. Barankay, and I. Rasul (2005). Social Preferences and the Response to Incentives: Evidence from Personnel Data. *The Quarterly Journal of Economics* 120(3), 917–962.
- Bandiera, O., I. Barankay, and I. Rasul (2007). Incentives for Managers and Inequality Among Workers: Evidence from a Firm-level Experiment. *The Quarterly Journal of Economics* 122(2), 729–773.
- 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.
- Beattie, G., R. Durante, B. Knight, and A. Sen (2017). Advertising Spending and Media Bias: Evidence from News Coverage of Car Safety Recalls. Technical report, National Bureau of Economic Research.
- Bénabou, R. and J. Tirole (2006). Incentives and Prosocial Behavior. *The American Economic Review* 96(5), 1652–1678.
- Bolton, P., M. Dewatripont, et al. (2005). *Contract theory*. MIT press.
- 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.
- Budak, C., S. Goel, and J. M. Rao (2016). Fair and balanced? quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly* 80(S1), 250–271.
- Butschek, S., P. Kampkötter, and D. Sliwka (2017). Paying gig workers. Working paper.
- Di Tella, R. and I. Franceschelli (2011). Government Advertising and Media Coverage of Corruption Scandals. *American Economic Journal: Applied Economics* 3(4), 119–151.
- Dohmen, T. and A. Falk (2011). Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender. *The American Economic Review* 101(2), 556–590.

- Dohmen, T. J., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2005). Individual risk attitudes: New evidence from a large, representative, experimentally-validated survey.
- Finan, F., B. A. Olken, and R. Pande (2015). The Personnel Economics of the State. Technical report, National Bureau of Economic Research.
- Freedom House (2017, April). Freedom of the Press Report, 2017. Technical report.
- Gentzkow, M. and J. M. Shapiro (2006). Media Bias and Reputation. *Journal of Political Economy* 114(2), 280–316.
- Gentzkow, M., J. M. Shapiro, and M. Sinkinson (2011). The Effect of Newspaper Entry and Exit on Electoral Politics. *American Economic Review* 101(7), 2980–3018.
- Hengel, E. (2016). Publishing While Female. Technical report, Technical report, University of Liverpool.
- Hoffman, M. and S. V. Burks (2017). Worker Overconfidence: Field Evidence and Implications for Employee Turnover and Returns from Training. Technical report, National Bureau of Economic Research.
- Holmstrom, B. and P. Milgrom (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica: Journal of the Econometric Society*, 303–328.
- 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.
- Lohner, J., S. Banjac, and I. Neverla (2016). Mapping structural conditions of journalism in Kenya. Technical report, MeCoDEM working paper.[Online]. Available from: <http://www.mecodem.eu/publications/working-papers>.
- McKenzie, D. (2012). Beyond baseline and follow-up: The case for more t in experiments. *Journal of development Economics* 99(2), 210–221.

- 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.
- Mohanan, M., G. Miller, K. Donato, Y. Truskinovsky, and M. Vera-Hernandez (2017). Different Strokes for Different Folks: Experimental Evidence on the Effectiveness of Input and Output Incentive Contracts for Health Care Providers with Different Levels of Skills.
- Mullainathan, S. and A. Shleifer (2005). The Market for News. *American Economic Review* 95(4), 1031–1053.
- Muralidharan, K. and V. Sundararaman (2011). Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy* 119(1), 39–77.
- Murtha, J. (2015). What It’s Like to Get Paid for Clicks. *Columbia Journalism Review*.
- Namwaya, O. (2018, February). Kenya: Government Crackdown Threatens Rights. Technical report.
- Oberholzer-Gee, F. and J. Waldfogel (2009). Media Markets and Localism: Does Local News en Espanol Boost Hispanic Voter Turnout? *American Economic Review* 99(5), 2120–28.
- 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. (2008). On the role of uncertainty in the risk-incentives tradeoff. *The BE Journal of Theoretical Economics* 8(1), 1–23.
- Schulhofer-Wohl, S. and M. Garrido (2013). Do Newspapers Matter? Short-run and Long-run Evidence from the Closure of the Cincinnati Post. *Journal of Media Economics* 26(2), 60–81.

- Shearer, B. (2004). Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment. *The Review of Economic Studies* 71(2), 513–534.
- Simiyu, T. F. (2014). Media Ownership and the Coverage of the 2013 General Election in Kenya—Democracy at the Crossroads. *Global Media Journal-African Edition* 8(1), 114–145.
- Snyder Jr, J. M. and D. Strömberg (2010). Press Coverage and Political Accountability. *Journal of Political Economy* 118(2), 355–408.
- Zabojnik, J. (1996). Pay-performance sensitivity and production uncertainty. *Economics Letters* 53(3), 291–296.

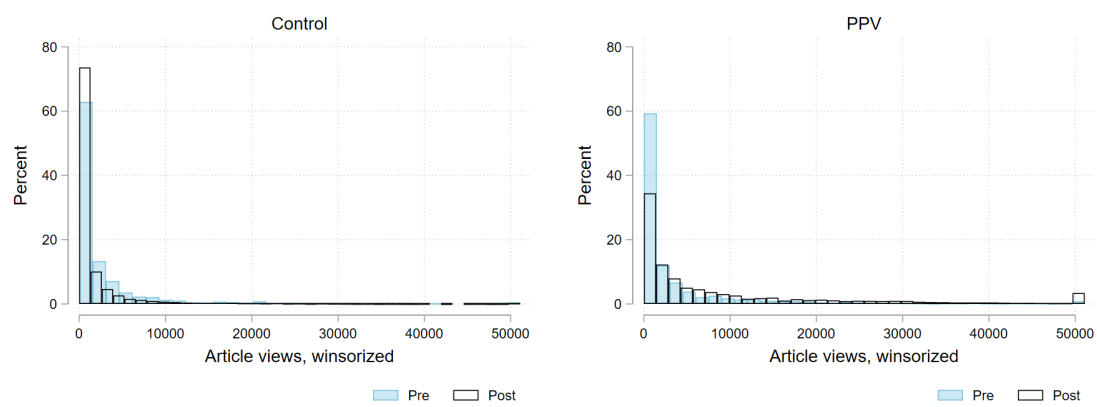


Figure 1: Distribution of pageviews, before and after treatment

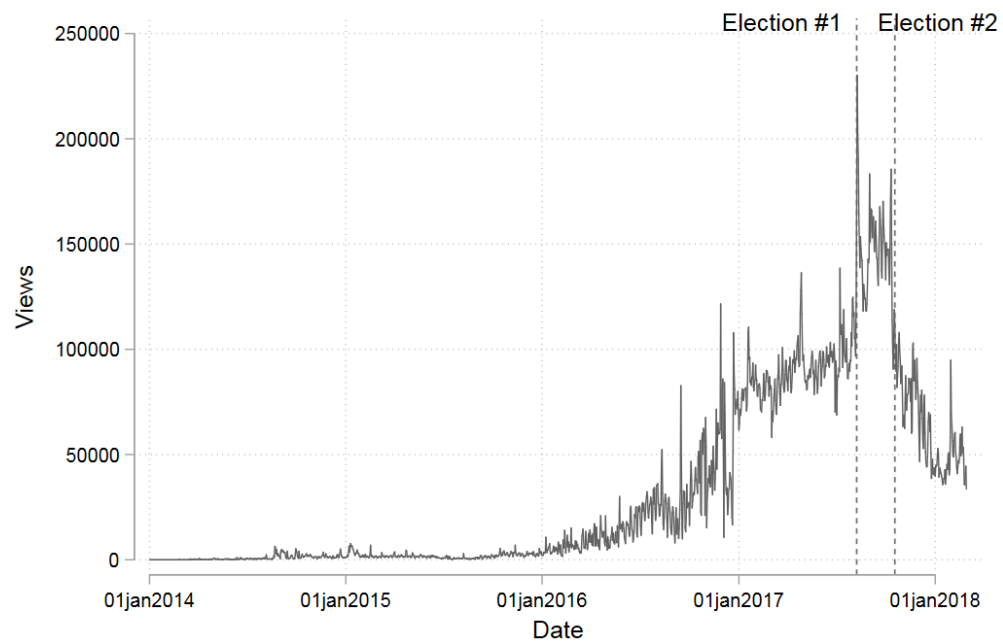


Figure 2: Daily pageviews – vertical lines mark the presidential elections

Article visitors	Your Pay
300	0
400	50
800	100
5000	150
15000	270
35000	510

Figure 3: Payment schedule

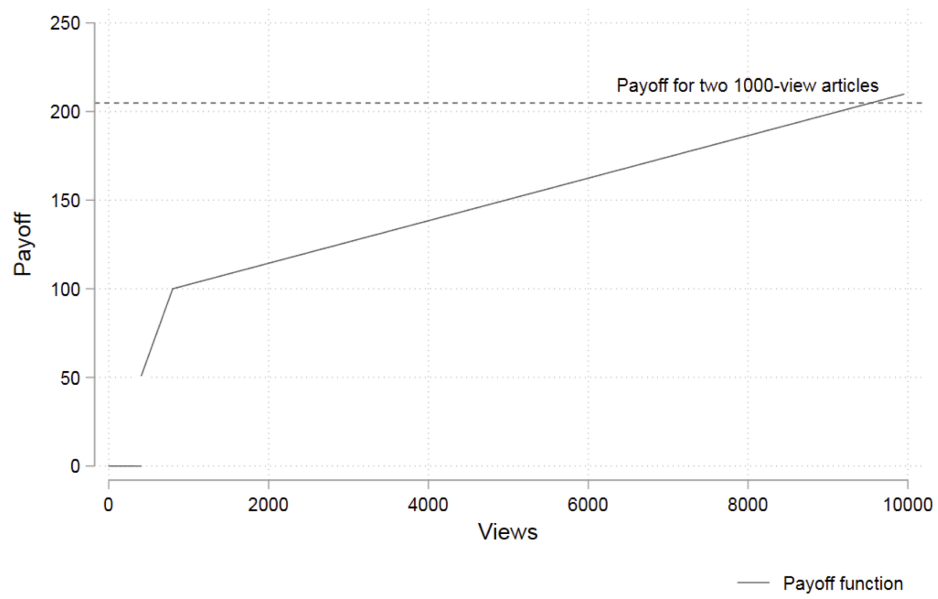


Figure 4: Payment schedule for PPV writers

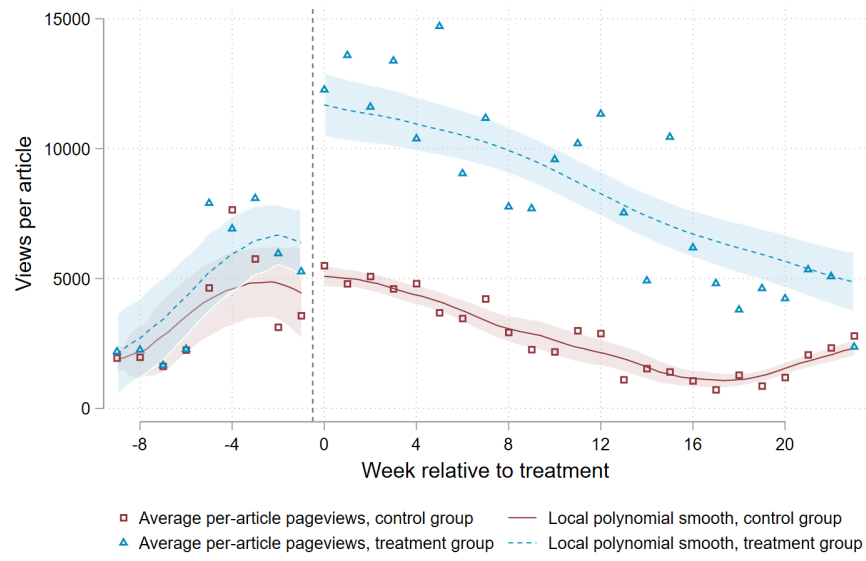


Figure 5: Average users per article, by treatment – weekly

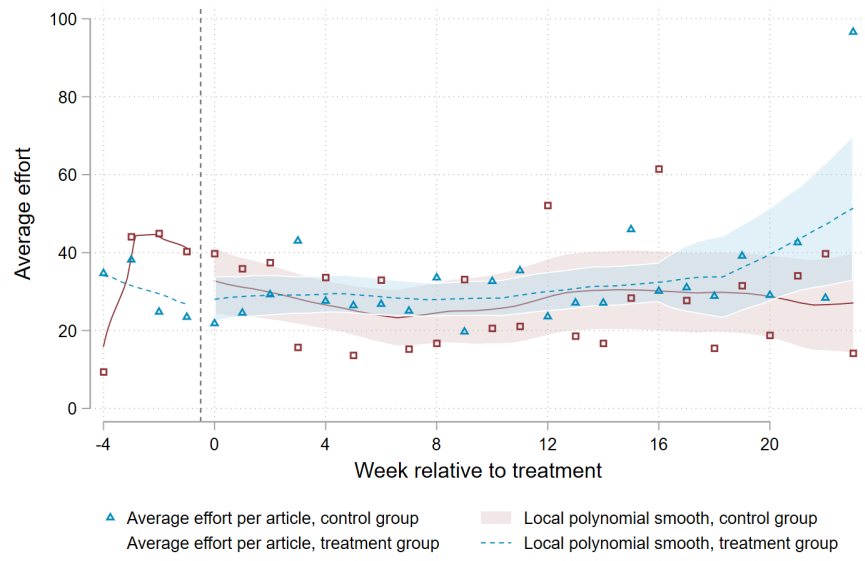


Figure 6: Average per-article effort - weekly

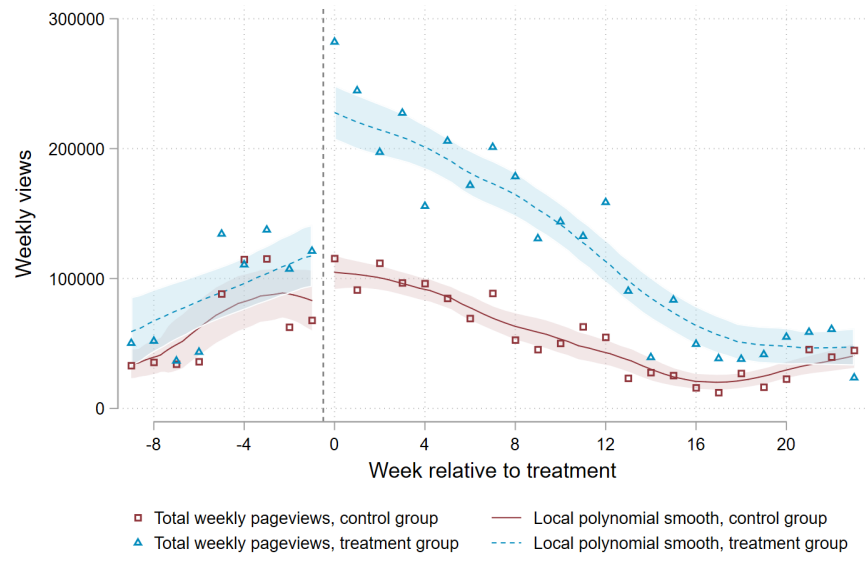


Figure 7: Total views by treatment - weekly

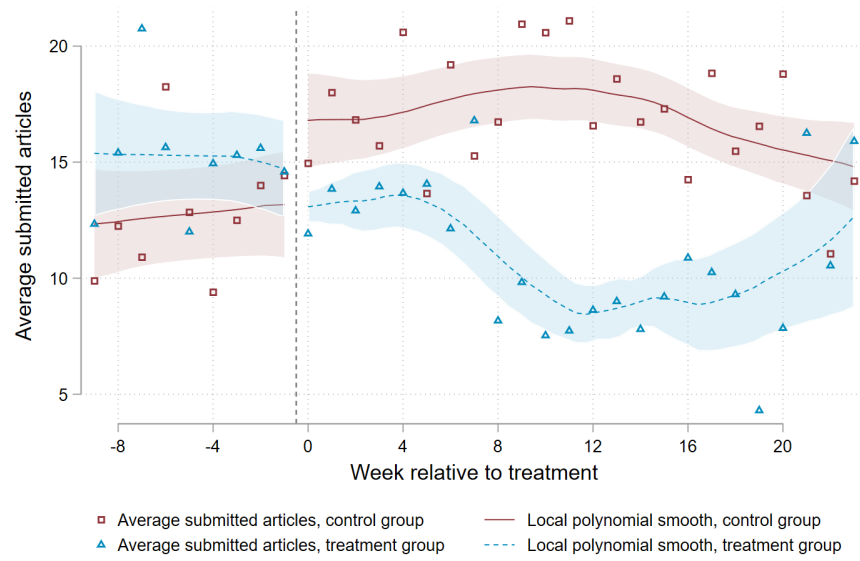


Figure 8: Average submitted articles by treatment - weekly

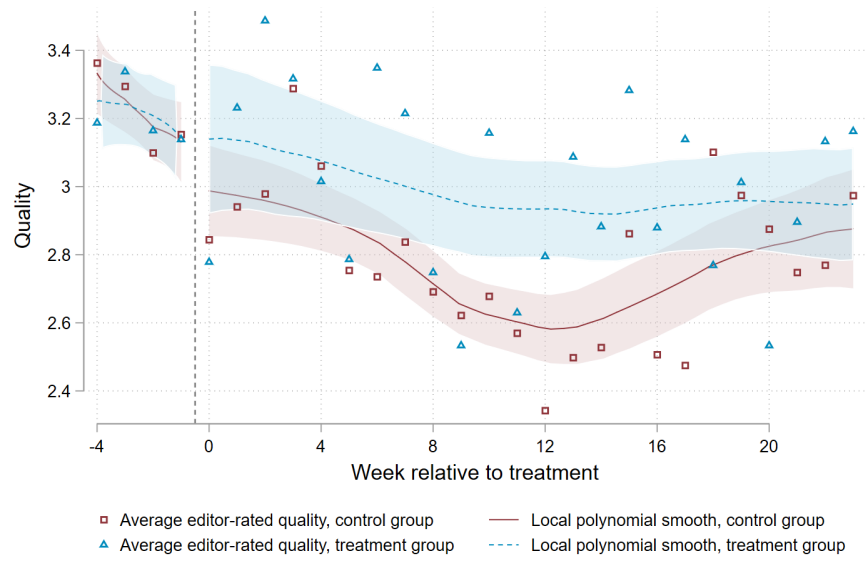


Figure 9: Average editor-rated quality by treatment - weekly

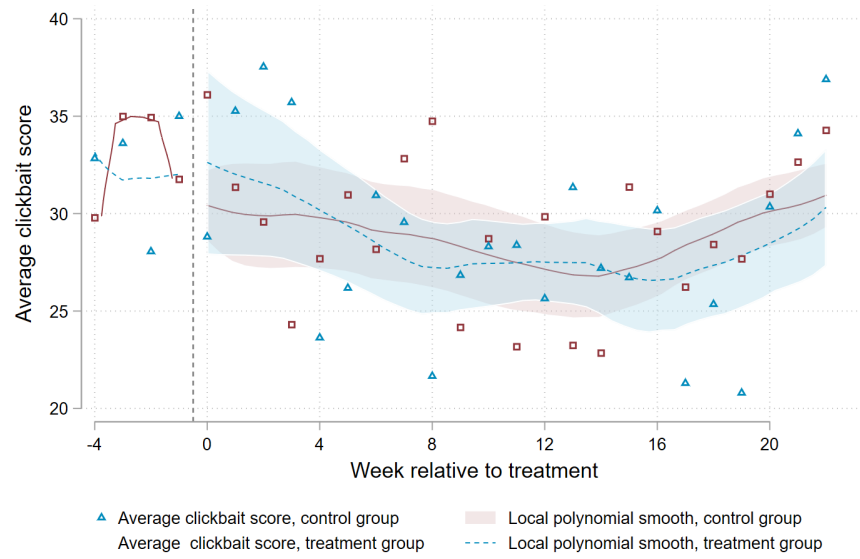


Figure 10: Average clickbait score by treatment - weekly

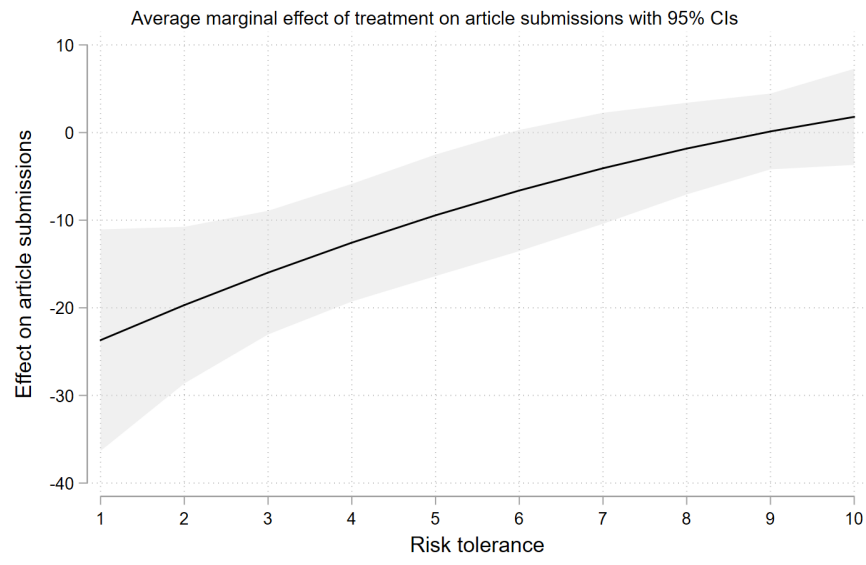


Figure 11: Treatment effects on article submissions by risk tolerance

Table 1: Balance at baseline across treatments

Variable	(1) Control		(2) PPV		(3) Choice		<i>p</i> -value from <i>t</i> -test Difference	
	N	Mean	N	Mean	N	Mean	(1)-(2)	(1)-(3)
	[strata]	(s.e.)	[strata]	(s.e.)	[strata]	(s.e.)		
Tenure	53 [14]	4.3 (1.3)	54 [14]	4.3 (1.4)	54 [16]	3.8 (1.1)	.92	.37
Submitted	53 [14]	16 (8)	54 [14]	20 (10)	54 [16]	17 (9.1)	.37	.85
Published	53 [14]	16 (8)	54 [14]	20 (10)	54 [16]	17 (9.1)	.37	.85
Views (month)	53 [14]	88,680 (58,500)	54 [14]	154,499 (108,543)	54 [16]	162,738 (116,580)	.24	.24
Views (article)	53 [14]	2,428 (1,319)	54 [14]	3,163 (1,560)	54 [16]	2,948 (1,495)	.18	.24
Share politics	53 [14]	.26 (.1)	54 [14]	.23 (.099)	54 [16]	.27 (.12)	.6	.79
Share county	53 [14]	.32 (.13)	54 [14]	.32 (.11)	54 [16]	.3 (.12)	.92	.78
Effort (article)	53 [14]	864 (487)	54 [14]	589 (346)	54 [16]	490 (269)	.56	.42
Effort (month)	53 [14]	864 (487)	54 [14]	589 (346)	54 [16]	490 (269)	.56	.42
Risk tolerance	30 [9]	6.9 (.46)	24 [7]	5.8 (.71)	30 [9]	7 (.4)	.27	.88
Cost/click	53 [14]	.03 (.012)	54 [14]	.02 (.0072)	54 [16]	.024 (.012)	.18	.22

Notes: The value displayed for *t*-tests are *p*-values. Standard errors are clustered at variable strata. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 2: ITT estimates of treatment on article impact

	Views/article	KES/view	Views	Views
PPV=1	4981.6*** (1355.0)	-0.087*** (0.020)	40339.5*** (13462.8)	44029.5*** (14030.9)
Tenure	99.8 (234.3)	-0.0076 (0.0053)	5142.5** (2258.3)	5331.2** (2354.2)
Constant	-547.8 (2586.9)	0.32*** (0.045)	-29631.8 (27127.6)	-34697.7 (29504.4)
Week FE	Yes	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes	Yes
Observations	803	803	873	803
Adjusted R^2	0.377	0.259	0.453	0.454
Mean dep. var	2805.9	0.15	33247.8	35930.3

Standard errors (in parentheses) clustered at the writer level.

Columns 1, 2, 4 include writer-weeks in which the writer published at least one article.

Column 3 includes all writer-weeks following treatment.

Excluded controls: indicator for whether writer was active in month prior to treatment and baseline weekly averages of outcome variable

Table 3: ITT estimates of treatment on writer effort

	Mins/article (sub)	Mins/article (pub)	Minutes/week	Minutes/week
PPV=1	1.07 (5.19)	2.37 (4.28)	-250.0** (115.0)	-250.3** (115.0)
Tenure	-4.50*** (1.66)	-3.85*** (1.19)	5.71 (14.8)	5.71 (14.8)
Constant	63.8*** (14.6)	56.8*** (10.9)	124.0 (134.0)	128.6 (135.0)
Week FE	Yes	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes	Yes
Observations	859	800	873	862
Adjusted R^2	0.142	0.175	0.267	0.265
Mean dep. var	37.3	35.6	612.7	626.3

Standard errors (in parentheses) clustered at the writer level.

Columns 1, 2, 4 include writer-weeks in which the writer published at least one article.

Column 3 includes all writer-weeks following treatment.

Excluded controls: indicator for whether writer was active in month prior to treatment and baseline weekly averages of outcome variable

Table 4: ITT estimates of treatment on writer submissions

	N Submitted	N Submitted	N Published	N Published
PPV=1	-7.26*** (2.36)	-7.45*** (2.51)	-6.22*** (2.10)	-6.39*** (2.24)
Tenure	0.99** (0.39)	1.03** (0.42)	0.98*** (0.37)	1.03** (0.39)
Constant	-0.44 (3.51)	-0.25 (3.98)	-2.32 (3.04)	-2.27 (3.42)
Week FE	Yes	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes	Yes
Observations	873	803	873	803
Adjusted R^2	0.357	0.324	0.362	0.327
Mean dep. var	16.6	17.8	14.2	15.4

Standard errors (in parentheses) clustered at the writer level.

Columns 1 & 3 include all writer-weeks following treatment.

Columns 2 & 4 include only writer-weeks in which the writer published at least one article.

Excluded controls: indicator for whether writer was active in month prior to treatment
and baseline weekly averages of outcome variable

Table 5: Treatment effects on writer submissions by risk aversion

	N Submitted	N Submitted	N Published	N Published
PPV=1	-25.3*** (5.73)	-28.0*** (9.13)	-23.8*** (5.74)	-23.7*** (8.90)
Risk tolerance	-1.29* (0.68)	-2.41 (2.47)	-1.30** (0.65)	-1.05 (2.36)
PPV=1 x Risk tolerance	2.83*** (0.71)	4.44 (3.37)	2.76*** (0.72)	2.60 (3.20)
Risk tolerance ²	0.11	(0.21)	-0.024	(0.20)
PPV=1 x Risk tolerance ²		-0.15 (0.27)		0.015 (0.26)
Tenure	0.91** (0.40)	0.95** (0.40)	0.87** (0.35)	0.86** (0.36)
Constant	12.0* (6.50)	12.8* (7.27)	10.2* (5.81)	10.0 (6.58)
Week FE	Yes	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes	Yes
Observations	799	799	799	799
Adjusted R^2	0.396	0.396	0.409	0.408
Mean dep. var	17.4	17.4	14.9	14.9

Standard errors (in parentheses) clustered at the writer level.

All columns include all writer-weeks following treatment.

Excluded controls: indicator for whether writer was active in month prior to treatment and baseline weekly averages of outcome variable

Table 6: ITT estimates of treatment on news content

	Quality	Clickbait	Share Political	Share County
PPV=1	0.018 (0.20)	-0.46 (1.33)	0.29*** (0.047)	-0.20*** (0.067)
Tenure	-0.13*** (0.038)	-1.26*** (0.43)	-0.027 (0.017)	0.033* (0.018)
Constant	2.84*** (0.34)	38.3*** (4.23)	0.15 (0.16)	0.66*** (0.18)
Week FE	Yes	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes	Yes
Observations	724	771	803	803
Adjusted R^2	0.326	0.101	0.438	0.388
Mean dep. var	2.78	29.0	0.35	0.61

Standard errors (in parentheses) clustered at the writer level.

Columns 1, 2, 4 include writer-weeks in which the writer published at least one article.

Column 3 includes all writer-weeks following treatment.

Excluded controls: indicator for whether writer was active in month prior to treatment and baseline weekly averages of outcome variable

Table 7: ITT estimates of treatment on article readability

	Syllables	words	Flesh	Dale-Chall	Coleman
PPV=1	70.4 (71.4)	48.3 (46.3)	-2.53 (1.87)	-0.32** (0.12)	-0.61* (0.31)
Tenure	-6.90 (12.2)	-4.84 (7.76)	-0.26 (0.59)	0.097* (0.050)	0.28*** (0.093)
Constant	370.0*** (99.4)	228.1*** (63.9)	49.9*** (5.04)	7.80*** (0.47)	13.0*** (0.90)
Week FE	Yes	Yes	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes	Yes	Yes
Observations	803	803	803	803	803
Adjusted R^2	0.070	0.054	0.237	0.243	0.282
Mean dep. var	511.7	334.7	66.0	9.13	14.6

Standard errors (in parentheses) clustered at the writer level.

All columns include writer-weeks in which the writer published at least one article.

Excluded controls: indicator for whether writer was active in month prior to treatment and baseline weekly averages of outcome variable

Table 8: ITT estimates of treatment on writer and reader behavior

	Ratio SN	Bounce	Time on page
PPV=1	0.010*** (0.0035)	-134.7 (134.6)	10.6 (6.93)
Tenure	0.00064 (0.0014)	19.3 (17.3)	2.40 (1.78)
Constant	0.0014 (0.015)	34.0 (212.4)	66.5*** (16.2)
Week FE	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes
Observations	803	803	803
Adjusted R^2	0.246	-0.015	0.441
Mean dep. var	0.019	346.4	150.9

Standard errors (in parentheses) clustered at the writer level.

Outcomes are (1) ratio of article views coming from social media,
(2) the rate at which readers visit articles w/o continuing to other pages,
and (3) the number of seconds spent reading the article.

All columns include writer-weeks in which the writer published ≥ 1 article.

Excluded controls: indicator for whether writer was active in month prior
to treatment, baseline weekly averages of outcome variable

Table 9: Determinants of opting in to the PPV contract

	Opt-in	Opt-in	Opt-in	Opt-in	Opt-in	Opt-in
Risk tolerance	0.35*** (0.11)	0.40*** (0.11)	0.43*** (0.10)	0.073*** (0.026)	0.072** (0.027)	0.077** (0.036)
Tenure		-0.15** (0.064)	-0.15** (0.076)		-0.048* (0.026)	-0.052 (0.032)
Avg. views $t - 4$			-0.000038 (0.00012)			0.0000049 (0.000020)
Avg. views $t - 3$			0.000048 (0.000064)			0.0000040 (0.000018)
Avg. views $t - 2$			-0.000059 (0.000097)			-0.00000038 (0.000035)
Avg. views $t - 1$			0.000036 (0.000051)			0.0000039 (0.000016)
Constant	-3.06*** (0.91)	-2.59*** (0.87)	-2.90*** (0.86)	-0.20 (0.17)	0.081 (0.26)	0.028 (0.29)
Strata FE	No	No	No	Yes	Yes	Yes
Observations	36	36	36	31	31	31
Adjusted R^2				0.124	0.190	0.034

Standard errors (in parentheses) clustered at the writer level.

Column 1 & 2 present marginal effects from a probit model, columns 3 & 4 from a linear probability model.

Table 10: Difference-in-differences with and without writer fixed effects

	Views/article	Views/article	Mins/article	Mins/article	Cost/view	Cost/view	N subm.	N subm.
PPV x post	5084.4*** (1431.6)	5901.2*** (1705.2)	15.8 (19.7)	21.3 (24.1)	-0.075*** (0.017)	-0.071*** (0.018)	-8.07*** (2.72)	-11.6*** (2.90)
PPV	990.9 (780.9)		-20.5 (21.7)		-0.027** (0.012)		3.42 (2.86)	
post	-723.4* (427.5)	-451.9 (555.2)	-21.7 (17.7)	-28.8 (22.7)	0.069*** (0.016)	0.063*** (0.018)	4.94** (2.27)	5.26** (2.19)
Tenure	-139.1 (177.5)		-2.40 (1.68)		-0.0029 (0.0041)		1.27*** (0.45)	
Constant	4672.6*** (1462.8)	3612.0*** (523.7)	76.1*** (20.1)	53.3*** (10.8)	0.098*** (0.033)	0.067*** (0.0074)	2.42 (3.80)	14.3*** (1.05)
Strata FE	Yes	No	Yes	No	Yes	No	Yes	No
Writer FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1125	1125	1060	1060	1125	1125	1206	1206
Adjusted R^2	0.256	0.474	0.010	0.090	0.176	0.243	0.171	0.518
Mean dep. var	2987.7	2987.7	41.4	41.4	0.13	0.13	16.1	16.1

Standard errors (in parentheses) clustered at the writer level.

Columns 1, 3, 5, 7 are difference-in-difference regressions.

Columns 2, 4, 6, 8 include writer fixed effects.

Appendix A: Proofs of Propositions 1 and 2

Proof of Proposition 1: Given a payment contract (α, β) , an agent with a coefficient of absolute risk aversion ρ solves

$$\max_{(n,e)} \mathbb{E}\{-\exp[-\rho(\alpha n + \beta en + \beta \varepsilon - c_1 ne^2 - c_2 n^2)]\}.$$

By a well known property of CARA utility functions²², the optimization problem is equivalent to solving

$$\max_{(n,e)} \alpha n + \beta en + \beta nm - c_1 ne^2 - c_2 n^2 - \frac{\rho}{2} \beta^2 n \sigma^2.$$

Under the flat contract, the agent does not benefit from choosing any $e > 0$ and her optimization problem can be simplified to:

$$\max_n n - c_2 n^2. \tag{FC}$$

This objective function is strictly concave and $n_{FC}^* = \frac{1}{2c_2}$.

Under the linear contract $(0, \beta)$, the agent solves

$$\max_{(n,e)} \beta en + \beta nm - c_1 ne^2 - c_2 n^2 - \frac{\rho}{2} \beta^2 n \sigma^2. \tag{LC}$$

Denote the objective function in (LC) by $F(n, e)$. The function F is concave in its arguments if and only if $\frac{\partial^2 F}{\partial n^2} \leq 0$ and

$$\frac{\partial^2 F}{\partial n^2} \frac{\partial^2 F}{\partial e^2} \geq \left(\frac{\partial^2 F}{\partial n \partial e} \right)^2.$$

²²Namely, we know that

$$\mathbb{E}[-\exp(-\rho x)] = -\exp(-\rho(\mu - (\rho/2)\sigma^2))$$

for any $x \sim N(\mu, \sigma^2)$. See for example [Bolton et al. \(2005\)](#).

The first inequality is verified readily ($-2c_2 < 0$). The second inequality simplifies to:

$$4c_1c_2n \geq (\beta - 2c_1e)^2.$$

It is immediate that this is true if and only if

$$\beta \in [2c_1e - 2\sqrt{c_1c_2n}, 2c_1e + 2\sqrt{c_1c_2n}]. \quad (3)$$

We next show that F must attain its maximum at some (n, e) at which (3) holds. Let

$$F^* := \max_{(n,e): \beta \in [2c_1e - 2\sqrt{c_1c_2n}, 2c_1e + 2\sqrt{c_1c_2n}]} F(n, e).$$

Assume that (3) does not hold; namely, let $\beta \notin [2c_1\tilde{e} - 2\sqrt{c_1c_2\tilde{n}}, 2c_1\tilde{e} + 2\sqrt{c_1c_2\tilde{n}}]$ for some (\tilde{n}, \tilde{e}) .

Setting $e^* = \frac{\beta}{2c_1}$, note that whenever $e = e^*$, (3) holds for all $n \geq 0$. This implies that $\tilde{e} \neq e^*$ and so:

$$\begin{aligned} F^* &\geq F(\tilde{n}, e^*) = \frac{\beta^2\tilde{n}}{2c_1} + \beta\tilde{n}m - \frac{\beta^2\tilde{n}}{4c_1} - c_2\tilde{n}^2 - \frac{\rho}{2}\beta^2\tilde{n}\sigma^2 \\ &= \frac{\beta^2\tilde{n}}{4c_1} + \beta\tilde{n}m - c_2\tilde{n}^2 - \frac{\rho}{2}\beta^2\tilde{n}\sigma^2 \\ &\geq \beta e\tilde{n} - c_1e^2\tilde{n} + \beta\tilde{n}m - c_2\tilde{n}^2 - \frac{\rho}{2}\beta^2\tilde{n}\sigma^2 \\ &= F(\tilde{n}, \tilde{e}), \end{aligned}$$

where the second inequality follows from the fact that for all $\tilde{e} \neq e^*$, $(c_1\tilde{e} - \beta/2)^2 > 0$ and so $\frac{\beta^2}{4c_1} > \beta\tilde{e} - c_1\tilde{e}^2$. The second inequality is also strict if $\tilde{n} > 0$.

Thus F must attain its maximum for some (n, e) for which (3) holds or, equivalently, F attains its maximum in the region over which it is concave. From the first-order condition, we get $e_{LC}^* = \frac{\beta}{2c_1}$ whenever $n^* > 0$ ²³ and so:

$$F(n, e_{LC}^*) = n \left(\frac{\beta^2}{4c_1} + \beta m - c_2n - \frac{\rho}{2}\beta^2\sigma^2 \right)$$

²³If $n^* = 0$, the value of e does not change the value of the objective function and e^* is indeterminate.

and this can be positive if and only if $\frac{\beta}{4c_1} + m - \frac{\rho}{2}\beta\sigma^2 > 0$. This, together with the first-order condition for n , gives us $n_{LC}^* = \max \left\{ 0, \frac{1}{2c_2} \left(\frac{\beta^2}{4c_1} + \beta m - \frac{\rho}{2}\beta^2\sigma^2 \right) \right\}$. \square

Proof of Proposition 2: Plugging in $\beta_{LC} = 1/m$ into the expressions for (n_{LC}^*, e_{LC}^*) from Proposition 1 gives

$$(n_{LC}^*, e_{LC}^*) = \left(\max \left\{ 0, \frac{1}{2c_2} \left(\frac{1}{4m^2c_1} + 1 - \frac{\rho\sigma^2}{2m^2} \right) \right\}, \frac{1}{2mc_1} \right).$$

It is sufficient to compare the value of

$$F((\alpha, \beta), (n, e)) = \alpha n + \beta en + \beta nm - c_1 ne^2 - c_2 n^2 - \frac{\rho}{2} \beta^2 n \sigma^2$$

evaluated at $((1, 0), (n_{FC}^*, e_{FC}^*))$ and at $((0, 1/m), (n_{LC}^*, e_{LC}^*))$. First, the maximum achievable utility of the agent under the flat contract does not depend on her risk aversion:

$$F((1, 0), (n_{FC}^*, e_{FC}^*)) = \frac{1}{4c_2} > 0.$$

Then:

$$\begin{aligned} & \lim_{\rho \rightarrow 0} F((0, 1/m), (n_{LC}^*, e_{LC}^*)) \\ &= \lim_{\rho \rightarrow 0} n_{LC}^* \left(\frac{1}{m} \frac{1}{2mc_1} + \frac{1}{m} m - c_1 \frac{1}{4m^2(c_1)^2} - c_2 n_{LC}^* - \frac{\rho}{2} \frac{1}{m^2} \sigma^2 \right) \\ &= \frac{1}{2c_2} \left(\frac{1}{4m^2c_1} + 1 \right) \left(\frac{1}{2m^2c_1} + 1 - \frac{1}{4m^2c_1} - \frac{1}{2} \left(\frac{1}{4m^2c_1} + 1 \right) \right) \\ &= \frac{1}{2c_2} \left(\frac{1}{4m^2c_1} + 1 \right) \left(\frac{1}{8m^2c_1} + \frac{1}{2} \right) \\ &> \frac{1}{4c_2} = F((1, 0), (n_{FC}^*, e_{FC}^*)). \end{aligned}$$

Therefore, for all sufficiently low $\rho \geq 0$, the agent prefers the linear over the flat contract.

Conversely:

$$\lim_{\rho \rightarrow \infty} F((0, 1/m), (n_{LC}^*, e_{LC}^*)) = 0 < F((1, 0), (n_{FC}^*, e_{FC}^*)),$$

because n_{LC}^* is zero for all sufficiently large ρ . So, for high levels of risk aversion, the agent prefers the flat over the linear contract.

Finally, we show that $F((0, 1/m), (n_{LC}^*, e_{LC}^*))$ is decreasing in ρ over the values of ρ , for which $n_{LC}^* > 0$:

$$\begin{aligned}
& \frac{\partial F((0, 1/m), (n_{LC}^*, e_{LC}^*))}{\partial \rho} \\
&= \frac{\partial n_{LC}^*}{\partial \rho} \left(\frac{1}{m} \frac{1}{2mc_1} + \frac{1}{m} m - c_1 \frac{1}{4m^2(c_1)^2} \right) - \frac{\partial n_{LC}^*}{\partial \rho} 2c_2 n_{LC}^* - \frac{1}{2} \frac{1}{m^2} n_{LC}^* \sigma^2 - \frac{\partial n_{LC}^*}{\partial \rho} \frac{\rho}{2} \frac{1}{m^2} \sigma^2 \\
&= \frac{\partial n_{LC}^*}{\partial \rho} \left(\frac{1}{4m^2 c_1} + 1 - 2c_2 \frac{1}{2c_2} \left(\frac{1}{4m^2 c_1} + 1 - \frac{\rho \sigma^2}{2m^2} \right) - \frac{\rho}{2} \frac{1}{m^2} \sigma^2 \right) - \frac{1}{2} \frac{1}{m^2} n_{LC}^* \sigma^2 \\
&= - \frac{1}{2} \frac{1}{m^2} n_{LC}^* \sigma^2 \\
&< 0.
\end{aligned}$$

This is sufficient to establish the existence of the critical value $\bar{\rho}$.

□

Appendix B: Main results estimated at the month-level

Table 11: ITT estimates of treatment on article impact

	Views/article	KES/view	Views	Views
PPV=1	4121.4*** (1267.3)	-0.067*** (0.018)	66422.4 (40315.9)	194213.9** (95742.4)
Baseline	0.25 (0.19)	0.95 (0.58)	-87211.9 (264297.7)	-405771.8 (822440.8)
Tenure	369.2 (294.7)	-0.014* (0.0073)	13729.2** (6595.0)	23031.4* (12559.1)
Constant	-2552.8 (3304.5)	0.36*** (0.068)	-852.6 (32901.6)	-67279.5 (110699.3)
Month FE	Yes	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes	Yes
Observations	279	279	642	279
Adjusted R^2	0.326	0.216	0.153	0.145
Mean dep. var	2937.6	0.14	53217.5	111336.5

Standard errors (in parentheses) clustered at the writer level.

Columns 1, 2, 4 include writer-months in which the writer published at least one article.

Column 3 includes all writer-months following treatment.

Excluded controls: indicator for whether writer was active in month prior to treatment.

Table 12: ITT estimates of treatment on writer effort

	Mins/article (sub)	Mins/article (pub)	Minutes/month	Minutes/month
PPV=1	2.54 (4.57)	3.48 (4.26)	-564.8** (224.0)	-881.7** (400.0)
Baseline	0.52*** (0.065)	0.53*** (0.062)	0.43*** (0.050)	0.43*** (0.057)
Tenure	-3.31*** (1.13)	-3.67*** (1.29)	32.9 (28.8)	57.4 (50.0)
Constant	52.8*** (8.69)	54.1*** (9.96)	259.6* (152.4)	415.4 (339.0)
Month FE	Yes	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes	Yes
Observations	313	279	642	313
Adjusted R^2	0.255	0.367	0.451	0.419
Mean dep. var	36.0	34.4	980.7	1845.4

Standard errors (in parentheses) clustered at the writer level.

Columns 1, 2, 4 include writer-months in which the writer published at least one article.

Column 3 includes all writer-months following treatment.

Excluded controls: indicator for whether writer was active in month prior to treatment.

Table 13: ITT estimates of treatment on writer submissions

	N Submitted	N Submitted	N Published	N Published
PPV=1	-18.0*** (5.26)	-29.1*** (8.73)	-15.2*** (4.71)	-24.6*** (7.60)
Baseline	0.55*** (0.11)	0.60*** (0.093)	0.53*** (0.098)	0.58*** (0.077)
Tenure	1.70** (0.77)	3.27** (1.51)	1.59** (0.69)	3.15** (1.39)
Constant	3.11 (3.85)	-2.79 (11.2)	1.36 (3.31)	-6.69 (9.86)
Month FE	Yes	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes	Yes
Observations	642	279	642	279
Adjusted R^2	0.409	0.361	0.404	0.362
Mean dep. var	26.5	55.2	22.7	47.6

Standard errors (in parentheses) clustered at the writer level.

Columns 1 & 3 include all writer-months following treatment.

Columns 2 & 4 include only writer-months in which the writer published at least one article.

Excluded controls: indicator for whether writer was active in month prior to treatment.