

Does Paying Referees Expedite Reviews?: Results of a Natural Experiment

Gary D. Thompson,* Satheesh V. Aradhyula,† George Frisvold,‡ and Russell Tronstad§

A natural experiment in an economics field journal afforded time-series observations on payments to referees for on-time reviews. The natural experiment yielded 15 months' worth of data with no payments and about two subsequent years of data with payments. Using referee- and manuscript-specific measures as covariates, hazard models were used to gauge the effects of payments on individual referee's review times. All models indicate statistically significant reductions in review times owing to referee payments. Reductions in review times translate into significant reductions in first-response time (FRT). Median FRT was reduced from 90 to 70 days, a 22% reduction in the presence of payments. With payments, only 1% of the FRTs exceeded six months; without payments, 16% of the FRTs exceeded six months.

JEL Classification: J44, C41, Q00

1. Introduction

Scientific journals have employed peer review to appraise the quality of manuscripts for over three centuries (Zuckerman and Merton 1971). Referees providing peer reviews generally do so voluntarily as part of their professional duties; they often receive no payment for reviewing manuscripts. Referees' rewards for timely, high-quality reviews are usually limited to nonpecuniary benefits, such as standing in good stead with editors and recognition as a referee for a prestigious journal.

When requesting peer reviews, journal editors place their trust in referees to provide high-quality, timely reviews. While editors can try to choose referees who will provide high-quality reviews, editors are able to provide few incentives to elicit careful, timely reviews or impose penalties on delinquent referees (Freedman 2000). Referees write and submit reports in virtual anonymity whether reviews are single or double blind. If referees shirk or lag in completing their reviews, they usually suffer little damage to their public reputation.

* Department of Agricultural and Resource Economics, University of Arizona, P.O. Box 210023, 1110 E. James Rogers Way, Tucson, AZ 85721-0023, USA; Phone 520-621-6249; E-mail garyt@ag.arizona.edu; corresponding author.

† Department of Agricultural and Resource Economics, University of Arizona, P.O. Box 210023, 1110 E. James Rogers Way, Tucson, AZ 85721-0023, USA; Phone 520-621-6260; E-mail satheesh@ag.arizona.edu.

‡ Department of Agricultural and Resource Economics, University of Arizona, P.O. Box 210023, 1110 E. James Rogers Way, Tucson, AZ 85721-0023, USA; Phone 520-621-6269; E-mail frisvold@ag.arizona.edu.

§ Department of Agricultural and Resource Economics, University of Arizona, P.O. Box 210023, 1110 E. James Rogers Way, Tucson, AZ 85721-0023, USA; Phone 520-621-2425; E-mail tronstad@ag.arizona.edu.

The authors would like to thank seminar participants at the Western Association of Agricultural Economics and the American Association of Agricultural Economics annual meetings in 2004 for valuable comments. We gratefully acknowledge the painstaking comments and suggestions of an anonymous referee. The manuscript benefited significantly from those suggestions.

Received January 2008; accepted February 2009.

Some economics journals offer payments as an incentive to referees for providing on-time reviews. Somewhat surprisingly, empirical research examining the efficacy of payments to encourage punctuality is quite limited. Apparently, only one empirical study has been conducted. Hamermesh (1991, 1994) provides cross-sectional evidence from a sample of 50 manuscripts submitted to seven journals in November 1989. Journals offering payments enjoyed a larger proportion of referee reports completed on time.

It is perhaps even more surprising that there is so little empirical research on the efficacy of referee payments given current review and publication lags at some economics journals. The increased time between the submissions of manuscripts to publication in economics journals over the last four decades has been documented thoroughly (Marshall 1959; Coe and Weinstock 1967; Yohe 1980; Laband, Maloney, and McCormick 1990; Trivedi 1993; Chong 2001; Ellison 2002). Ellison attributes about one quarter of the increase in the submission-to-publication lag to increased times between submission and the first editorial response, the first-response time (FRT). Azar (2004) argues that delays in FRT actually account for more of the total lag if FRTs of manuscripts previously rejected by journals other than the publishing journal are taken into account. Most journals offering referee payments do so only in the first round of reviews in an effort to reduce FRT.

The possible benefits of the slowdown in FRT to authors, referees, editors, and journal readers have been examined extensively by Azar (2005a, b, 2007, 2008). Most authors prefer quicker FRTs, and reductions in FRTs should lead to quicker dissemination of scientific results. However, shorter FRTs may encourage submission of lower-quality manuscripts to high-quality journals, resulting in costs to referees and editors. With more manuscripts to handle, rejection rates may climb, and authors may have to submit their manuscripts to more journals before manuscripts are ultimately accepted for publication. This article makes no judgment about the social welfare of reducing FRTs.¹ Rather, it tackles the more modest question of whether payments induce referees to expedite their reviews and thereby reduce FRTs.

Previous theoretical models for paying referees distinguish the participation rate—the rate at which referees agree to review manuscripts—from the review time. Chang and Lai (2001) find possible multiple equilibria in which payments may be offered by some journals because not all journals are prestigious enough to attract referees without compensation. Engers and Gans (1998) endogenize both the participation rates and the review time in considering the equilibrium level of payments. The equilibrium payment level is zero because neither the referees nor the editors fully internalize the effects of their decisions on journal quality. Hamermesh's (1991) appointment-book model also treats the participation rates and review times separately. Although the model does not unambiguously generate hypotheses regarding refusal rates and review times, the role of payments to referees is to elicit “queue-jumping,” that is, to influence the order in which voluntary tasks are completed by referees. These theoretical models highlight a salient feature of the market for referee services: Because referees provide their services voluntarily, there is no pricing mechanism to allocate effort.

Many economists would expect the answer to the question of whether payments to referees for on-time reviews can expedite first-round reviews to be self-evident: Payment for what was

¹ Payments may elicit quicker reviews, but they could result in reviews of poorer quality if referees prepare them hastily in order to meet deadlines. The referee's task of producing a review has at least two dimensions: punctuality and quality. If quality is poorly measured, it may not be socially optimal for a principal (editor) to pay a referee (an agent performing a task with several dimensions) for on-time reviews (Holmstrom and Milgrom 1991). We would like to thank an anonymous referee for calling our attention to this point.

previously a voluntary activity should not perversely affect effort and performance (Arrow 1972).² The only question might plausibly be the degree to which payments can expedite reviews. However, some empirical evidence suggests that when voluntary economic activities—giving blood, volunteering to work for public or private institutions, and collecting donations for charity, for example—are rewarded with relatively low payment levels, low-paid performance is *inferior* to voluntary performance. In their analysis of a cross-section of Israeli high school students collecting donations for charity, Gneezy and Rustichini (2000) find statistically significantly smaller donations collected by students to whom relatively low payments were offered; students volunteering and students receiving higher payments collected higher levels of donations. Econometric evidence from a cross section of workers in the Swiss Labor Force Survey indicates that relatively low levels of financial compensation for voluntary activities result in fewer hours volunteered per month compared to hours volunteered by workers who received either no compensation or higher levels of compensation (Frey and Goette 1999). These two studies suggest that paying for what was once a voluntary activity like refereeing manuscripts could produce unexpected results, especially if the payment is low. Different types of payments may elicit different effects on previously unpaid or voluntary tasks. In the two studies mentioned, levels of payment varied discretely (Gneezy and Rustichini 2000) or continuously (Frey and Goette 1999). In the natural experiment considered here, a single payment level was established. Whether a single payment level can expedite first-round reviews remains an open question.

The purpose of this study is to provide some empirical evidence about the effects of payments to referees for a single field journal, the *Journal of Agricultural and Resource Economics (JARE)*, through time. The question addressed is as follows: Do payments significantly reduce first-response times (FRT)? We have information on these response times both before and after payments were implemented at *JARE*. Hence, we can complement Hamermesh's (1994) cross-sectional comparisons of journals with and without payments by making time-series comparisons of the efficacy of payments to referees of the same journal. No new theory to explain on-time behavior is offered. Instead, new data from a natural experiment and econometric analysis assessing the impacts of payments on review and FRTs are presented.

2. The Natural Experiment

The natural experiment generating the data consists of editorial records for *JARE* before payment to referees and after implementing payments. *JARE* is the journal of the Western Association of Agricultural Economics (WAEA). The editorial duties of *JARE* typically rotate every three years. An editorial team of an editor and three co-editors assumed responsibility for *JARE* from April 2000 to April 2003. The editorial period of our data begins April 2000. For purposes of continuity, all initial submissions received through March 2003 were handled by the same editors until the first editorial response. All subsequent editorial correspondence and revised manuscripts were forwarded to the new editors, whose duties began in April 2003.

² A payment for what was a voluntary activity might be perceived by a referee as a means for an editor to force the referee to meet a minimum standard, namely, to produce a review by the deadline. In this case, the referee could perceive the incentive as a signal of distrust. Experimental evidence in Falk and Kosfeld (2006) suggests that controls imposed by the principal requiring the agent to meet minimum standards can result in poorer performance if the agent perceives the control as a signal of distrust.

In July 2001, the executive council of WAEA, the governing body responsible for approving new uses of the WAEA budget, voted to approve a payment of \$50 for on-time first-round reviews of manuscripts submitted to *JARE*. On-time reviews were taken to be those completed within six weeks of the referee agreeing to the editor's request to review a manuscript. This natural experiment generated 15 months' worth of reviews with no payment—April 2000 through June 2001—and about 24 months' worth of reviews under the payment regime. All reviews were double blind.

The protocol for soliciting reviews, whether with or without payments, was to send prospective referees an electronic mail message soliciting a review. After the referee agreed to review the manuscript, a copy of the manuscript with cover letter was sent either electronically or by regular mail to the referee. Although *JARE* policy encouraged electronic submission, many authors submitted manuscripts by mail. Manuscripts not submitted electronically were usually sent by courier to referees residing outside of North America to reduce lag times in foreign mail delivery. Nearly all referees' reports were submitted by electronic mail to the editors, facilitating exact dating of reports. In the majority of cases, the determination of an on-time submission of the referee's report was clear-cut because reports were submitted before the deadline. A one-day grace period for on-time reviews was allowed.

The editors of *JARE* almost always used two referees for each manuscript. In extenuating circumstances, a single referee report was used. In a few cases, a third referee report might have been used, even though it was received after the deadline. The natural experiment does not include data on participation rates before and during payments. Although individual editors kept track of participation rates for their own purposes of obtaining two referees per manuscript, participation rates were not systematically recorded. Hence, the natural experiment does not afford a measure of the impact of payments on participation rates.

The natural experiment provides data having some desirable characteristics. The within-journal temporal comparison holds editors, editorial policy with the exception of payments, editorial procedures, and the potential pool of referees constant across comparisons. In contrast, Hamermesh's study (1994) compared response times across journals, each of which had possibly different editorial policies, procedures, and pools of referees. The data from the natural experiment are not left-censored because there is no carryover of first reviews from the editor prior to April 2000. Right-censoring occurs only because a few referees failed to complete their reviews. No other journal in agricultural and resource economics, domestic or international, paid referees during the three-year sample period from 2000 to 2003. Hence, there are no external effects of another, potentially competing, field journal having decided to pay referees during the natural experiment.

The natural experiment has a few shortcomings. Participation rates were not systematically recorded. The start date for the review may have differed slightly from the date on which the manuscript was received. For the few manuscripts sent by regular mail to domestic referees or by courier to referees outside of North America, the receipt date slightly lags the start date.

3. Covariates and Descriptive Statistics

The natural experiment yielded covariates specific to referees and authors as well as covariates specific to manuscripts and referee reports. Each set of covariates is described in the following subsections.

Referee- and Author-Specific Characteristics

Payments were not designed to compensate referees for time spent reviewing manuscripts. However, one might expect referees with higher incomes to be less responsive to a payment. Instead of attempting to measure income, we use two proxies for income: years of experience since earning a Ph.D. and number of publications listed in *EconLit*. In order to separate the effects of time in service from productivity, we constructed a *per annum* measure of publications based on publications mentioned in *EconLit*. Hamermesh, Johnson, and Weisbrod (1982) find years in service and academic citations to be the two most significant determinants of academic salaries.

Gender is another variable used in some empirical studies (Hamermesh 1991; Ellison 2002). We also include an indicator variable for female referees but have no *a priori* expectations about gender differences in refereeing behavior. Neither Hamermesh (1991) nor Ellison (2002) find statistically significant results by gender.

Affiliation with top-ranked departments was measured in order to compare potential differences across referees and authors. Although all reviews were double blind, authors from top-ranked schools might tend to write better and more easily evaluated manuscripts because they benefit from colleagues' reviews from within their departments. Rankings of departments are usually drawn from two sources: surveys of opinions and indices of citations of faculty (Dusansky and Vernon 1998). Although reputation as measured by surveys may not track contemporaneously with publication productivity as measured by citations, Dusansky and Vernon (1998) find relatively high correlations among the ranks of economics departments based on these disparate sources. For purposes of ranking agricultural economics departments in the United States and Canada, we used several rankings generated from citations (Tauer and Tauer 1984; Beilock, Polopolous, and Correal 1986; Beilock and Polopolous 1988) and a more recent survey by Perry (1994). We find positive correlations among the rankings of these studies. An unweighted linear combination of the rankings from Perry's survey with the rankings from Beilock and Polopolous' *Social Science Citations Index* per capita measure was chosen. The departments included in the top 20 rankings were relatively stable across alternative ranking schemes.³

Authors and referees residing outside the United States and Canada were identified by an indicator variable. Hamermesh (1994) posits that reviewers affiliated with institutions outside the United States would likely be less committed to United States-based journals and would, therefore, be less inclined to submit prompt reviews. Even with electronic mail and courier delivery, foreign referees are likely subject to more lag time in receiving manuscripts. Foreign academicians also may face different incentives: research and publishing activities may be rewarded differently than at some U.S. and Canadian universities.

Employment of a referee by the U.S. federal government was noted. Federal law prohibits federal employees from receiving remuneration of any sort for professional duties like refereeing.

An indicator variable measuring whether the referee was also the author of a manuscript under review in *JARE* at the same time was constructed. Referees whose work was currently

³ Procedures used for combining rankings are available on request from the authors. The top 20 departments in order were as follows: University of California (UC) at Berkeley, Iowa State, Minnesota, Stanford, Maryland, UC Davis, Wisconsin, Illinois, Cornell, North Carolina State, Michigan State, Purdue, Oklahoma State, Florida, Ohio State, Oregon State, Pennsylvania State, Texas A&M, Virginia Tech, and Georgia.

under review would be more likely to return reviews promptly in order to build goodwill with editors.

Referees from the same department as that of the editors were also identified. These in-house referees might respond more quickly to editors for various reasons. Editors can make requests and reminders face to face. In-house referees may also want to preserve or enhance collegiality by providing on-time reports.

Manuscript-Specific Characteristics

Nearly all empirical studies of publication lags find that longer manuscripts take longer to review (Laband, Maloney, and McCormick 1990; Hamermesh 1991, 1994; Chong 2001; Ellison 2002). Accordingly, we include the number of pages as a measure of the length of the manuscript. We wanted to include a measure of complexity or difficulty of the manuscript because more technically demanding manuscripts usually take longer to review. However, we could not find an unambiguously defensible measure of complexity or difficulty.

The number of co-authors is measured to assess the impact of more authors on review times. More authors suggest potential for more coordination problems among authors with the possible deterioration in the coherence and uniformity of quality throughout the manuscript. Ellison (2002) presents conflicting evidence regarding the effect of number of authors: In the 1970s, it was associated with a reduction in publication lags, but there was a positive association in the 1990s.

The ordinal rank of the corresponding author—whether listed as first author, second author, etc.—is measured to control for possible influence on review times. Laband (2002) finds strong circumstantial evidence that order of authorship matters critically among agricultural economists, so we expect that the ordinal rank of the corresponding author might influence review time through quality of preparation of the manuscript. The further down the list of authors is the corresponding author, the less likely it is that the corresponding author has exerted substantial quality control over the manuscript. The resulting poorer-quality manuscripts may take more time to peruse and critique.

Review-Specific Characteristics

Time to completion of review is measured by the number of days from the date the referee agreed to review the manuscript to the date when the completed referee report was received. We also include a measure of size of the review. Not all referee reports were received electronically, making a word count tedious and subject to error. Instead of word count, we rely on a less precise but indicative measure, namely, the number of pages of single-spaced prose, including equations and figures. Finally, we record the incidence of referee reports never completed, which results in right censoring of the time to completion.

Descriptive Statistics

Variable definitions are provided in Table 1, and descriptive statistics for referees and authors are displayed in Table 2. The sample of corresponding authors and manuscripts is about one half the size of the sample of referees and referee report statistics because two

Table 1. Variable Definitions

Variable Name	Definition
Referee characteristics	
<i>Years experience</i>	Integer number of years from the year in which the referee's Ph.D. was granted to the year in which the referee report was requested. If Ph.D. had not yet been granted, value of zero was assigned.
# <i>Articles EconLit</i>	Number of articles attributed to the referee, regardless of number of coauthors or order of authorship, cited in <i>EconLit</i> . Article count does not include working papers. No time limit was imposed on the past years for the number of articles cited.
# <i>Articles EconLit per annum</i>	# <i>Articles EconLit</i> divided by <i>Years experience</i> .
# <i>Articles JARE</i>	Number of articles attributed to the referee, regardless of number of coauthors or order of authorship, in the <i>Journal of Agricultural and Resource Economics (JARE)</i> .
# <i>Citations JARE</i>	Number of times the referee's <i>JARE</i> articles were cited as listed in <i>Social Science Citation Index</i> .
<i>Female</i>	Dichotomous variable for gender; equals one if female.
<i>Top 20 school</i>	Dichotomous variable for current referee affiliation with a department of agricultural and resource economics ranked in the top 20; equals one if a member of top 20 department when the review was requested.
<i>Foreign</i>	Dichotomous variable for referee residing outside of United States and Canada; equals one if not in North America.
<i>Federal employee</i>	Dichotomous variable for a referee employed by the U.S. federal government; equals one if a federal employee. Federal employees may not accept monetary or in-kind compensation for professional duties such as refereeing.
<i>Also author</i>	Dichotomous variable for a referee who was also an author of a manuscript under review at the <i>Journal of Agricultural and Resource Economics</i> at the time the referee report was requested. The manuscript could have been in any stage of review but not yet accepted for publication.
<i>Referee same dept.</i>	Dichotomous variable for a referee who was a faculty member of the same academic department as the editors at the time the referee report was requested.
Manuscript characteristics	
# <i>Pages of manuscript</i>	Integer number of pages of text not including title page, abstract, references, tables, or figures.
# <i>Coauthors</i>	Number of coauthors excluding the corresponding author.
<i>Rank corresponding author</i>	The ordinal rank of the corresponding author. No adjustment for equal authorship was made.

referees were generally chosen for each manuscript. The respective samples are divided into the period of about 15 months before payments were implemented ("Before") and approximately 24 months during which payments were made ("During").

There are only a few systematic differences in the sample statistics before and during payments. Most notably, average days to completion of review are fewer during the payment period. The size of referee reports also tended to be larger during the payment period. The

Table 2. Sample Means, Before and During Payments to Referees

Category	Before	During	<i>p</i> value
Review characteristics			
Days to complete review	49.6	41.6	0.002
Number of pages of review	1.5	1.7	0.024
% of reviews never completed	3.6%	2.1%	0.322
Referee characteristics			
Years experience	14.2	14.4	0.827
Number articles <i>EconLit</i>	18.3	21.7	0.127
Number articles <i>EconLit per annum</i>	1.48	1.67	0.183
Number articles <i>JARE</i>	1.51	1.66	0.382
Number citations <i>JARE</i>	2.2	2.1	0.797
Female	10.7%	10.4%	0.994
Top 20 school	42.9%	43.2%	0.914
Foreign	2.7%	5.3%	0.137
Federal employee	7.6%	4.5%	0.099
Also author	9.8%	11.2%	0.639
Referee same department	1.8%	2.9%	0.372
<i>Sample size (# of referees)</i>	224	375	
Manuscript characteristics			
Number of pages	18.3	18.1	0.779
Number of coauthors	1.60	1.57	0.827
Rank of corresponding author	1.28	1.32	0.558
Corresponding author characteristics			
Years experience	9.5	10.6	0.316
Number of articles in <i>EconLit</i>	14.6	16.5	0.464
Number of articles in <i>EconLit per annum</i>	1.6	1.5	0.462
Number of articles in <i>JARE</i>	1.339	1.873	0.057
Number of citations in <i>JARE</i>	1.147	1.868	0.084
Female	15.6%	18.0%	0.599
Top 20 school	28.4%	38.1%	0.092
Nonuniversity affiliated	12.8%	9.0%	0.317
Federal employee	11.9%	5.3%	0.061
Foreign	3.7%	5.8%	0.388
<i>Sample size (# of manuscripts)</i>	109	189	

“Days to complete review” excludes right-censored observations, which occurred when delinquent referees never completed their reviews. Boldfaced entries denote statistically significant differences in means at the 10% level using a two-tailed test.

incidence of incomplete reviews fell when payments were made, but the difference is not statistically significant.

On average, the pool of referees appears comparable before and during payments. Likewise, the sample of authors differed little across the two periods.

4. Econometric Evidence

The process by which referees agree to review and complete reviews could be modeled as a sequential decision in which the binary participation decision conditions subsequent time to completion of the review. Although we recognize censoring exists due to the participation

decision, we do not have participation data with which to model that decision. Thus, we focus only on the behavior of referees once they have agreed to review a manuscript.

Even after agreeing to review manuscripts, not all referees returned reviews. The failure of delinquent referees to finish reviews resulted in right censoring of the dependent variable, *Days to Complete Review*. The incidence of right censoring is relatively small—16 of 599 reviews—so the effects on alternative estimators are likely modest, even when censoring is accounted for in estimation.

A subset of the covariates may be variation in time throughout the sample period. For single manuscript reviews, we assume that referee-specific characteristics that can change through time—years of experience, professorial rank, institutional affiliation, number of citations, etc.—are fixed during the review period because those characteristics change infrequently. However, for referees who reviewed more than one manuscript during the three-year period, some of these referee-specific characteristics did change from one review to the next. We use time-varying values of these covariates in estimating the hazard models.

In addition to the referee- and manuscript-specific covariates included in the duration model, an indicator variable for the period in which payments were made is used to capture any systematic differences associated with payments. Several parametric and semiparametric estimators of the duration models were considered to assess the robustness of payment effects across specifications. A Weibull duration model, which permits positive, constant, or negative duration dependence, was estimated, as was a Weibull–gamma mixture model to accommodate possible unobserved heterogeneity. Score tests of Jaggia and Trivedi (1994) led to rejection of the null of no unobserved heterogeneity and duration dependence. A flexible duration model—the ordered logit model of Han and Hausman (1990)—was also used because it allows baseline hazards to be estimated without imposing arbitrary parametric forms. Additionally, we applied Lewbel’s ordered data model, which extends the Han and Hausman model in the presence of heteroskedasticity of unknown form (Lewbel 2000; Stewart 2005).⁴

Each of the foregoing models was applied using the full sample of referees, but some of the referees reviewed manuscripts before payments were made and subsequently with payments. The subsample of repeat referees allows us to match repeat referee review times with and without payments, thereby controlling for unobserved heterogeneity. The sample size using only repeat referees is less than half that of the full sample: 236 observations on repeat referees versus 599 for all referees. As selection of repeat referees is not random, we apply a selectivity–duration model (Prieger 2002) in which the selectivity equation is specified as a probit, and the duration equation is specified as a Weibull model.

Tests for fixed effects across editors—*JARE* had one editor and three co-editors—were performed, but no statistical differences were found. Differences in the effects of payments across the last two years of the sample were also performed. Evidence of payments having statistically different effects across the two years only occurred in the ordered logit–gamma mixture model.

For brevity, we report only the results from the ordered logit and selectivity–duration models.⁵ Although the focus of this analysis is on the efficacy of payments to elicit timely

⁴ In Lewbel’s estimator, “ordered” refers to a sorting of the residuals from the first-stage estimate for use in the second stage (for details, see Lewbel 2000, pp. 157–8).

⁵ Parameter estimates for all models applied to the full sample and the subsample of repeat referees are available from the authors. For the ordered logit model, parameter estimates of the weekly dichotomous variables are omitted owing to space constraints. Parameter estimates of the selection equation in the selectivity–duration model are omitted for the same reason.

Table 3. Selected Hazard Rate Models

Covariate	Full Sample	Repeat Refs
Constant	3.645 (0.000) ^a	3.018 (0.000)
Referee characteristics		
Years experience	-0.002 (0.824)	0.011 (0.072)
# articles <i>EconLit</i>	- 0.107 (0.084)	0.013 (0.686)
# articles <i>JARE</i>	- 0.076 (0.059)	- 0.032 (0.022)
# citations <i>JARE</i>	0.022 (0.238)	0.004 (0.524)
Female	0.333 (0.222)	0.227 (0.078)
Top 20 school	0.264 (0.113)	0.088 (0.257)
Foreign	-0.495 (0.275)	-0.021 (0.956)
Federal employee	-0.200 (0.538)	0.474 (0.156)
Also author	- 0.511 (0.057)	- 0.238 (0.029)
Referee same dept.	-0.052 (0.914)	-0.100 (0.613)
Manuscript characteristics		
# pages of manuscript	0.021 (0.227)	0.026 (0.005)
# coauthors	-0.026 (0.733)	0.001 (0.969)
Rank of corresponding author	0.017 (0.891)	0.042 (0.493)
Payment		- 0.194 (0.009)
Payment, year 1	- 0.638 (0.003)	
Payment, year 2	- 0.327 (0.094)	
Theta ^b	1.151 (0.075)	0.669 (0.000)
Omega ^c		- 0.839 (0.000)
Sample sizes	599	599/236
Log likelihood	-1378.7	-1433.0

^a p values for two-tailed tests are in parentheses. Estimated coefficients with $p < 0.100$ are in boldface type. Standard errors were calculated using the Hessian matrix of the respective log-likelihood functions.

^b Theta is the scale parameter of the gamma distribution.

^c Omega is the correlation parameter from Prieger's selectivity-duration model. The selection equation is specified as a probit and the duration portion as a Weibull. Other specifications of the selectivity and duration equations produced qualitatively similar results.

referee reports, the effects of the other covariates on duration time are briefly examined. A few of the referee-specific characteristics tend to have statistically detectable effects.⁶ Most notably, referees with a manuscript under review tended to complete their reports more quickly. Also, referees having published more articles in *JARE* tended to finish their reports more rapidly. In the subsample of repeat referees, longer manuscripts led to longer review times.

The dichotomous variables for referee payments had significant and negative impacts on review times in all models estimated, not just the two models reported in Table 3. The reduction in mean and median review times was generally about one week. In the full sample, mean and median review times without payments were just under seven weeks; whereas in the subsample of repeat referees, the mean and median review times were about five weeks without payments. The reductions of mean or median review times are statistically significant in all cases.

To appreciate the impacts of payments on expediting review times, we plot the cumulative distribution functions (CDFs) for completing reviews from the ordered logit and selectivity-duration models. With the ordered logit, the probabilities of completing a review in a given week may be estimated. Figure 1 displays the CDFs with and without payments. In weeks four

⁶ The magnitudes of parameter estimates are not comparable between the ordered logit and selectivity-duration models. Accordingly, only the sign and statistical significance of the parameter estimates are discussed.

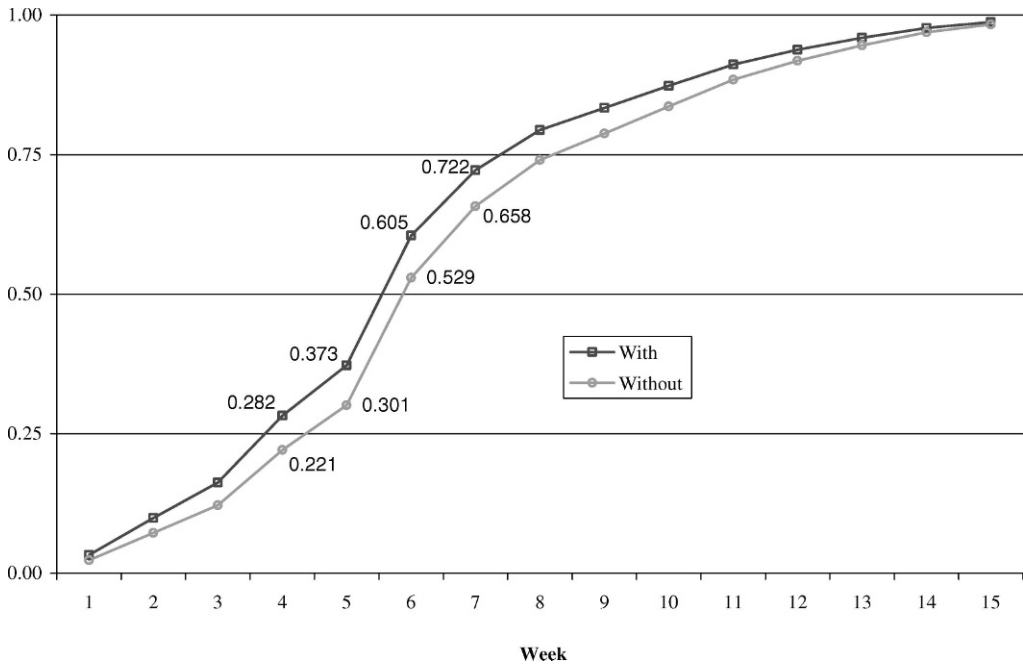


Figure 1. Cumulative distribution of probabilities of completing review by week, all referees, with and without payments, ordered logit model

through seven, the cumulative probabilities of completing reviews with payments exceed those without payments by more than 0.060. The maximum difference in cumulative probabilities of 0.075 occurs in week six, the deadline for on-time reviews. Payments result in a CDF for completing reviews lying everywhere to the left of the CDF without payments, indicating notable increases in the likelihood of early or on-time reviews.

Figure 2 depicts the CDF of expected review times derived from the probit–Weibull model for repeat referees. The expected review times appear short because most repeat referees reviewed manuscripts more quickly than nonrepeaters. The impact of payments is striking: Expected review times are reduced with payments for nearly all review times. For example, median review times were reduced by 7 days from 37 days without payment, a 19% reduction in time. With payments, 99% of reviews were completed in 44 days, versus 53 days without payments, a 17% reduction in time. Payments exert noteworthy effects on both slow and quick repeat referees.

5. Payment Effects on First-Response Times

The natural experiment provides ample econometric evidence that payments expedite individual referee’s reports, whether in the full sample or the subsample of repeat referees. However, for the majority of manuscripts, two referee reports were sought. Hence, the effects of payments on FRTs depend in most cases on how quickly pairs of referee reports are completed. To appreciate the effects of payments on FRTs, consider Figure 3, in which the empirical cumulative distributions (CDF) of FRTs with and without payments are plotted. The CDF of

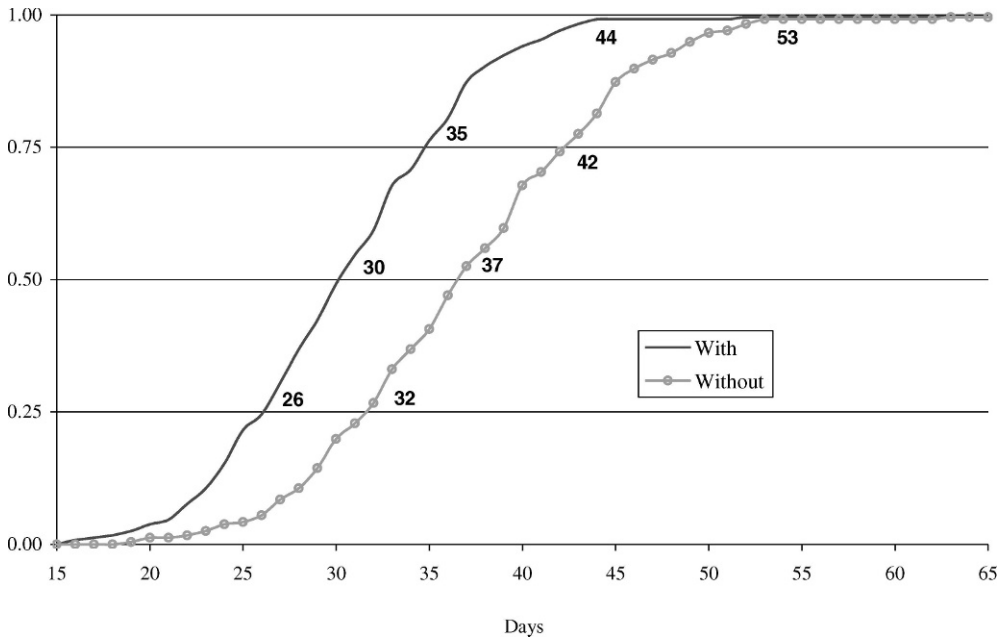


Figure 2. Cumulative distributions of day to complete reviews, repeat referees, with and without payments, Probit-Weibull model

FRTs with payments lies entirely to the left of the CDF without payments, indicating payments appreciably reduce FRTs when referees are paired. The change in median FRTs from 90 to 70 days represents a 22% reduction. At the 75th percentile, there is nearly a 32% reduction in FRT from 138 to 94 days. Put differently, with payments, three quarters of the manuscripts had

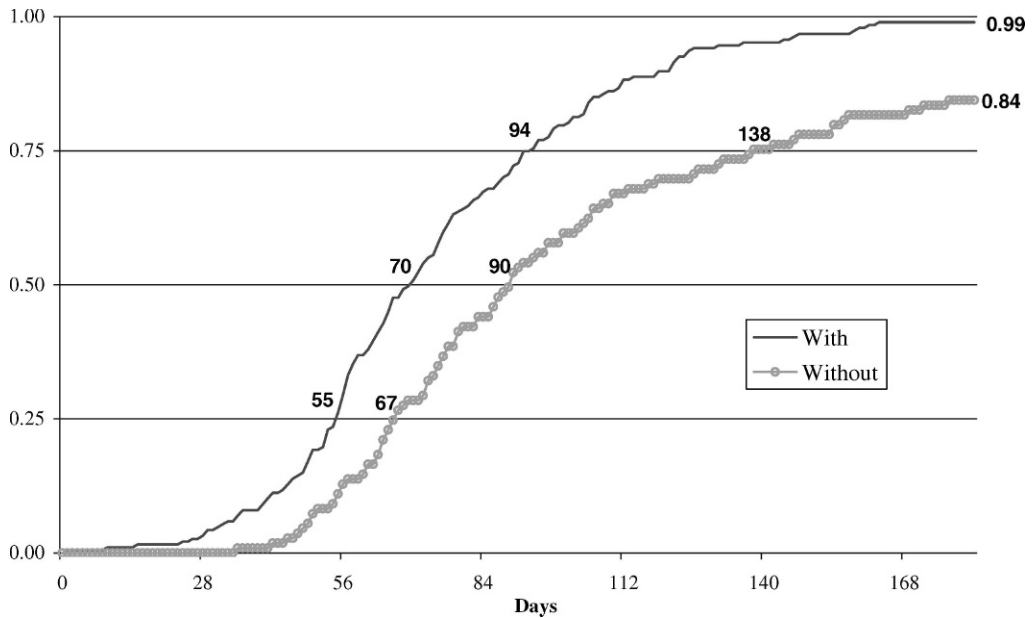


Figure 3. Cumulative Distributions of First-Response Time, with and without Payments

FRTs of about 13 weeks or less; whereas without payments, only half of the manuscripts had been processed in about the same time. Also, with payments, virtually all manuscripts had FRTs of less than half a year; whereas without payments, only 84% of the manuscripts had a first editorial response in less than six months.

FRTs are composed of three sequential periods: (i) the time from receipt of the manuscript to when each referee agrees to review, denoted as t_1^i , $i=1,2$; (ii) the referee's review time, t_2^i ; and (iii) the editor's handling of the manuscript after receiving the referee's report, t_3^i . Start dates for each of these three periods will only coincide by happenstance, so FRT depends on the timing and size of t_k^i , $i=1,2$; $k=1,2,3$. Payments provide an incentive designed to shorten t_2^i to at least the deadline, \bar{t}_2 . The effect of payments on t_1^i is not entirely observed in the sample because participation rates were not recorded. Payments influence t_3^i only indirectly. Should both referees finish before the deadline, $\max(t_2^1, t_2^2) < \bar{t}_2$, editors might not rush their editorial decision but could still attain a reasonable FRT. If payments do not induce either referee to complete timely reports, $t_2^i > \bar{t}_2 \forall i$, diligent editors will expedite their editorial decision to compensate for the tardy reviews. A single punctual referee would not necessarily allow for a short FRT if paired with a tardy review, $t_2^i < \bar{t}_2 < t_2^j$, $i \neq j$. The foregoing suggests that the impacts of payments to individual referees on shortening FRT could be attenuated by the pairing of referees and editors' behavior. Despite the potential for attenuation, in this natural experiment, payments substantially reduce FRTs.

6. Conclusions

Time-series evidence from a natural experiment on refereeing manuscripts at the *Journal of Agricultural and Resource Economics (JARE)* from 2000 to 2003 suggests that payments to referees can expedite review and FRTs. Referees did not dash off shorter reports to meet the deadline for payment; in fact, reports were statistically significantly longer with payments than they were prior to payments. Measuring the quality of referee reports before and during payments is difficult. Thorough referee reports are necessarily longer; of course, longer reports are not necessarily better. Yet the evidence suggests there was no decline in length of referee reports when payments were implemented. The percentage of delinquent referees declined with payments, though the difference was not statistically significant.

JARE editorial policy usually required two referee reports, so the effects of payments on reducing FRT depend on how payments might induce pairs of referees to respond on time. Although payments are designed to provide incentives to individual referees, when referees are paired in the natural experiment, FRT is reduced substantially in the presence of payments. Median FRT fell from 90 to 70 days, a 22% reduction. With payments, 99% of all manuscripts had FRT of less than half a year; whereas prior to payment, only 84% had FRT of six months or less.

The single-journal evidence presented here complements Hamermesh's cross-sectional study of seven economics journals in 1989 (Hamermesh 1994). One advantage of the single-journal data used here is that the potential pool of referees and authors is virtually homogeneous through time. The results for the subsample of repeat referees compare the *same* referees before and during payments, thereby reducing the effects of unobserved heterogeneity. Hamermesh's cross-sectional data compared referees from four general and three subspecialty

journals in economics. Hamermesh (1991) notes the anonymous journal “G1” employed higher-quality referees. One disadvantage of the single-journal data is that they do not control for editorial activities and policies from other journals. No other comparable field journals in agricultural and resource economics offered referee payments during the three years analyzed. Another disadvantage of single-journal data is that they may not indicate the effects of payments on referees who often review for general economics or other subspecialty journals that pay for refereeing manuscripts. Despite these shortcomings, the empirical evidence of the natural experiment underscores the efficacy of paying referees to expedite FRTs.

References

- Arrow, Kenneth J. 1972. Gifts and exchange. *Philosophy and Public Affairs* 1:343–62.
- Azar, Ofer H. 2004. Rejections and the importance of first response times. *International Journal of Social Economics* 31:259–74.
- Azar, Ofer H. 2005a. The review process in economics: Is it too fast? *Southern Economic Journal* 72:482–91.
- Azar, Ofer H. 2005b. The academic review process: How can we make it more efficient? *American Economist* 50:37–50.
- Azar, Ofer H. 2007. The slowdown in first-response times of economics journals: Can it be beneficial? *Economic Inquiry* 45:179–87.
- Azar, Ofer H. 2008. Evolution of social norms with heterogeneous preferences: A general model and an application to the academic review process. *Journal of Economic Behavior & Organization* 65:420–35.
- Beilock, Richard, and Leo Polopolous. 1988. Ranking of agricultural economics departments: Influence of regional journals, joint authorship, and self-citations. *American Journal of Agricultural Economics* 70:403–09.
- Beilock, Richard, Leo Polopolous, and Mario Correal. 1986. Ranking of agricultural economics departments by citations. *American Journal of Agricultural Economics* 68:595–604.
- Chang, Jui-jen, and Ching-chong Lai. 2001. Is it worthwhile to pay referees? *Southern Economic Journal* 68:457–63.
- Chong, Alberto. 2001. On the lags between submission and acceptance: Are all referees created equal. *Applied Economic Letters* 8:423–25.
- Coe, Robert K., and Irwin Weinstock. 1967. Editorial policies of major economic journals. *Quarterly Review of Economics and Business* 7:37–43.
- Dusansky, Richard, and Clayton J. Vernon. 1998. Rankings of U.S. economics departments. *Journal of Economic Perspectives* 12:157–70.
- Ellison, Glenn. 2002. The slowdown of the economics publishing process. *Journal of Political Economy* 110:947–93.
- Engers, Maxim, and Joshua S. Gans. 1998. Why referees are not paid (enough). *American Economic Review* 88:1341–49.
- Falk, Armin, and Michael Kosfeld. 2006. The hidden costs of control. *American Economic Review* 96:1611–30.
- Freedman, Craig. 2000. Do economic journals obey economic prescriptions? *Review of Industrial Organization* 17:371–84.
- Frey, Bruno S., and Lorenz Goette. 1999. Does pay motivate volunteers? Institute for Empirical Research in Economics Working Paper No. 7, University of Zürich, Zürich, Switzerland.
- Gneezy, Uri, and Aldo Rustichini. 2000. Pay enough or don't pay at all. *Quarterly Journal of Economics* 115:791–810.
- Hamermesh, Daniel S. 1991. The appointment-book problems and commitment, with applications to referring and medicine. NBER Working Paper No. 3928.
- Hamermesh, Daniel S. 1994. Facts and myths about refereeing. *Journal of Economic Perspectives* 8:153–63.
- Hamermesh, Daniel S., George E. Johnson, and Burton A. Weisbrod. 1982. Scholarship, citations, and salaries: Economic rewards in economics. *Southern Economic Journal* 49:472–81.
- Han, Aaron, and Jerry A. Hausman. 1990. Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics* 5:1–28.
- Holmstrom, Bengt, and Paul Milgrom. 1991. Multitask principal-agent analysis: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization* 7:24–52.
- Jaggia, Sanjiv, and Pravin K. Trivedi. 1994. Joint and separate score test for state dependence and unobserved heterogeneity. *Journal of Econometrics* 60:273–91.
- Laband, David N. 2002. Contribution, attribution, and the allocation of intellectual property rights: Economics versus agricultural economics. *Labour Economics* 9:125–31.
- Laband, David N., Michael T. Maloney, and Robert E. McCormick. 1990. The review process in economics: Some empirical findings. *Review of Economics and Statistics* 72A:5–17.

- Lewbel, Arthur. 2000. Semiparametric qualitative response model estimation with unknown heteroskedasticity or instrumental variables. *Journal of Econometrics* 97:145–77.
- Marshall, Howard D. 1959. Publication policies of the economics journals. *American Economic Review* 49:133–38.
- Perry, Gregory M. 1994. Ranking M.S. and Ph.D. graduate programs in agricultural economics. *Review of Agricultural Economics* 16:333–40.
- Prieger, James E. 2002. A flexible parametric selection model for non-normal data with application to health care usage. *Journal of Applied Econometrics* 17:367–92.
- Stewart, Mark B. 2005. A comparison of semiparametric estimators for the ordered response model. *Computational Statistics & Data Analysis* 49:555–73.
- Tauer, Loren W., and Janelle R. Tauer. 1984. Ranking doctoral programs by journal contributions of recent graduates. *American Journal of Agricultural Economics* 66:170–72.
- Trivedi, P. K. 1993. An analysis of publication lags in econometrics. *Journal of Applied Econometrics* 8:93–100.
- Yohe, Gary W. 1980. Current publication lags in economics journals. *Journal of Economic Literature* 18:1050–55.
- Zuckerman, Harriet, and Robert K. Merton. 1971. Patterns of evaluation in science: Institutionalisation, structure, and functions of the referee system. *Minerva* 9:66–100.