

(Un)Available upon Request: Field Experiment on Researchers' Willingness to Share Supplementary Materials

Michal Krawczyk, Ph.D.,¹ and Ernesto Reuben, Ph.D.²

¹Faculty of Economic Sciences, University of Warsaw, Warsaw, Poland

²Columbia Business School, New York, New York, USA

This article reports results of a field experiment in which two hundred e-mails were sent to authors of recent articles in economics that had promised to send the interested reader supplementary material, such as alternative econometric specifications, “upon request.” The e-mails were sent either by a researcher affiliated at Columbia University, New York or the University of Warsaw, Poland; furthermore, the authors' position (assistant professor) was specified in half the e-mails only. Overall, 64% of the approached authors responded to our message, of which two thirds (44% of the entire sample) delivered the requested materials. The frequency and speed of responding and delivering were very weakly affected by the position and affiliation of the sender. Gender or affiliation of the author, number of citations or journal impact factor or the type of object in question seemed to make no difference. However, authors of published articles were much more likely to share than authors of working papers.

Keywords: data sharing, field experiment

1. INTRODUCTION

There is broad consensus in the community that sharing of data, codes, stimuli and other supplementary research materials fosters progress of science. Indeed, the benefits are numerous. Perhaps most importantly, unnecessary duplication of effort is avoided. Replications are made possible, enabling identification of mistakes and fraud attempts. New methods and theories may be applied to existing data sets. Reviews and meta-analyses facilitate aggregation of knowledge. Young scholars may use others' materials learning research methodology and data analysis as well as treat them as models for their own

Address correspondence to Michal Krawczyk, Faculty of Economic Sciences, University of Warsaw, Warsaw, Poland. E-mail: mkrawczyk@wne.uw.edu.pl

projects. Science historians may track changes of research methods, etc. From an ethical viewpoint sharing is thus generally to be praised, although several limitations may be identified (Smith, 2011).

It is much more debatable whether sharing supplementary research materials actually pays off, especially in the short run. Feigenbaum and Levy (1993) and others argue that voluntary sharing will not work. Indeed, making data, additional results, or stimuli available will generally require some effort. Possibility of verification is a mixed blessing. Acceleration of others' projects may interfere with the author's own publication plans. These and similar reasons may explain why 75% of pharmaceutical researchers generally oppose raw data sharing (Kirwan et al., 1997).

As a natural response to such a social dilemma, some journals require making supplementary materials available at the time of publication. In economics, such decisions were fueled by spectacularly unsuccessful replication attempts undertaken by Dewald et al. (1986), McCullough and Vinod (2003), and others. Although Freese (2007) praises economists for their high standards of data sharing, few journals have actually begun to require archiving thus far and for those that have, enforcement proves to be difficult. McCullough et al. (2006) found that of 193 empirical papers published in the *Journal of Money, Credit, and Banking* that should have had entries in the journal's data repository, only 69 actually did, of which just 14 contained data and code permitting replication. Similar results hold for the Federal Reserve Bank of St. Louis Review and seemingly for most other journals that have established their archives more recently (McCullough et al., 2008). More generally, there is a concern that mandatory use of data repositories may backfire by encouraging some authors to choose another outlet, to postpone publication until all research plans associated with the used code or data set are concluded or perhaps even to avoid projects with high "fixed cost" altogether.

In the most common situation of the choice being effectively left to the author, the best strategy may well be to make a non-binding promise to share the supplementary materials and hope that nobody calls. In this way, the author may seize nearly all the benefits of sharing (signal compliance with norms, enhance her credibility, thus increase chances of publication, trigger positive reciprocity) with only some risk of bearing the costs (time of preparation, risk of flaws being discovered, etc.). Indeed, the commitment to send the data, proof of theorem, alternative econometric specifications, etc., "upon request" is repeatedly encountered in scholarly economic papers. The question that we tackle in this study is whether such a promise is in fact credible.

To that end, we sent requests to authors who have made explicit vows to share supplementary materials.

Our approach is most closely related to previous projects involving requests to share data sets as mandated by journal policy. These generally indicate a major compliance problem. For example, Wicherts et al. (2006) investigated

141 recent articles published in APA journals (that require signing Ethical Principles concerning i.a. sharing of data for re-analysis). They found that only 25.7% of datasets were sent within six months, despite repeated requests. In a follow-up analysis (Wichert et al., 2011), it was recently found that authors of studies with internal inconsistency in reported statistics were especially unlikely to share their data. Savage and Vickers (2009) only received one out of ten requested data sets, on which publications in two PLoS medical journals were based, despite these journals' explicit data sharing requirements and underlying "open access" approach.

Very little is known about the determinants of compliance with requests to share supplementary materials, such as the professional status of the person who made it. Freese (2007), in particular, suggests that Ph.D. students may have greater difficulty obtaining requested data sets than professors. The location and prestige of the institution from which the request comes is likely to affect (speed of) the response. For example, Trimble and Ceja (2011) show and cite some evidence that scholarly journals process submissions from U.S.-based authors quicker. The same may be true in the case of requests to share supplementary materials. At the same time, bibliometric studies (Glänzel and Schubert, 2005) show that scientific collaboration (e.g., co-authorship) is strongly affected by geographical distance, for example European scholars tend to work relatively more with other Europeans. Thus requests coming from a less distant location may fare better. This study is, to the best of our knowledge, the first one that systematically varies academic position and the home institution of the requester. Overall, we find that compliance rates are somewhat higher than in most previous studies, that promises made in published papers are more frequently honored than those in working papers, and that the position and prestige seem to make a small difference.

2. SAMPLE

Our sampling population consisted of all 2009 academic papers in economics and business listed on scholar.google that contained the phrase: "available upon request" or "from the author(s) upon request." This was a simple way to select papers in which some additional materials were indeed promised.

The choice of the year was, admittedly, somewhat arbitrary. On the one hand, we did not want to include older papers in our sample, for contact details would likely be outdated, the materials in question could easily be lost in the meantime or would not be available in a suitable digital form, etc. On the other hand, opting for very recent papers would make it more difficult to judge the quality of the study. In our sample it will be proxied by the impact factor of the journal and the number of citations. Of papers that appeared in, say, 2011, a larger fraction are still working papers; many have very few or no citations yet. For the same reason, we expected more rejections of data requests related

to very recent papers because the authors could still be working on extensions, related projects, etc.

Two papers on each results page of scholar.google (every fifth paper) were sampled. If the paper was not suitable (we could not access the full text, it did not seem to belong to the field of economic sciences or was not an academic publication at all, no additional material was in fact promised, the additional material was clearly available in a working paper version on-line, the nature of the promised material could not be grasped and described quickly, or we personally knew the authors), then the following paper on the same results page was used instead. Papers were also replaced whenever it was necessary to assure that any author (including non-corresponding ones) does not appear more than once in our database.

In each selected paper, the specific passage containing the authors' promise was identified. The object in question was precisely described and located (e.g., it could be "parameter estimates as well as their corresponding bootstrap standard errors for the other thirteen equations (footnote 7)"). It was also classified as belonging to one of the following categories: econometric results, data, data description, theoretical result/proof, stimuli (instructions, etc.), appendix/extension.

The first author was addressed, unless somebody else was indicated as corresponding author. Affiliations and e-mail addresses of the author were retrieved from the paper or otherwise found on the authors' personal websites. The primary affiliation was also coded as American (including Canada), European (including Israel), or other (typically Asian or Australian). Gender of the author was coded, whenever this could be done quickly and with high degree of confidence.

A typical e-mail could read:

"Dear Professor Robalo, I have come across your paper "Using Infinitely-Dimensional Smirnov Space to Analyze Small-Sample Properties of Quasi-Semi-Parametric Rao-Simao Estimators" and I was wondering whether you would be so kind to send me the estimates of the CCCSFA-GARCCH-X model for alternative assumptions about the values of kappa, deppa and pippa, as generously offered in the paper (footnote 125). This could be very useful for my research project.

Thank you very much for considering my request. Best regards
name, affiliation"

The e-mails were either sent from the account of Ernesto Reuben, University of Columbia or Michal Krawczyk, University of Warsaw. In either case, the sender was either signed as Assistant Professor or not. Fifty e-mails were sent in each of these four treatments.

The follow-up varied by case, depending on the response. Some authors simply sent the requested materials and some promised to do so at some point in the future. In either case, they were sent an appropriate thank-you note.

Some authors were not sure about the details of the request, in which case we provided such additional information. If they asked about the purpose, we simply repeated a very general statement about the research project, promising to acknowledge their help.

Unlike in most previous related studies that were by-products of search for specific data (e.g., for a meta-analysis), we did not really need the actual additional materials (only were trying to find out whether the author would be willing to send them or not).¹ Therefore, in the few cases in which the author mentioned it would cost a lot of time or effort to prepare/retrieve the requested material, we would send an e-mail explaining that we actually did not need it so badly. We want to emphasize that we do not think that our procedure involved substantial ethical concerns, as we only asked the authors to comply with their (publicly made) promise to perform an action dictated by the rules of proper conduct of research. However, for obvious reasons, it was impossible to obtain informed consent from our participants, thus we chose to stay on the safe side, whenever nontrivial effort on their part was signaled.² The study was approved by the Ethics Committee of the first author's institution; in particular, the Committee did not require that we debrief the subjects.

3. PREDICTIONS

Clearly, academic integrity requires that one delivers on promises made in scientific reports, especially when it is likely to facilitate further research. We therefore expected that most authors would be willing to help. On the other hand, as suggested in the introductory section, to the extent that fulfilling the request involves certain effort or risk, it may be that researchers are most willing to do so if the e-mail comes from a faculty member of a respected institution. Columbia University is currently number 8 at the Academic Ranking of World Universities, while Warsaw shares ranks 301–400. Although these figures, as at least one of us feels, may overstate the difference in academic excellence to some extent, they yield a clear prediction of higher response rate to e-mails sent from Columbia. We also expect more replies to the e-mails signed by a faculty member. Further, as long as strategic concerns, such as possibility of future cooperation, come into play, as well as solidarity within the broad community of U.S. or E.U. scholars, we expect more positive response on part of American (European) authors to the requests coming from the American (European) sender. We may also expect that response and compliance will tend to be faster whenever the request comes from the preferred sender.

Concerning the impact of such paper-specific variables as impact factor, number of citations or type of the requested material, there is little theory or previous evidence to go by. For example, authors of acclaimed, high-quality studies may have higher research standards, be less afraid of replication

attempts or be more likely to have already prepared the requested material for another reader. On the other hand, authors of less prestigious publications may be more eager to give it a chance finding a collaborator in the field.

4. RESULTS

4.1. Descriptive Statistics

Our sample covered a variety of journals. Naturally, outlets specialized in publishing short texts were over-represented, e.g., *Applied Economics Letters* had nine entries and *Economics Letters* had five. Mean five-year JCR-2010 impact factor for the 137 entries that had any was 1.07, and on average each paper had 7.48 scholar.google citations. JEL codes were copied whenever available. Eighty-six out of 200 authors had their primary affiliation at a European institution and 90 at an American one; about three-quarters were male. Of the objects in question, nearly 60% were econometric results, the rest spread more or less evenly over other categories.

For each author approached, we have coded whether they responded at all to our e-mail and if so, how many days it took. Similarly, we recorded whether and when the entire requested material was sent.³ Table 1 shows response and compliance rates, by treatment.

4.2. Treatment Effects

It turns out that response rates were slightly higher when sender identified himself as faculty member ($p = .039$ in the one-tailed test of proportions), although there was no significant difference in compliance rate. Conversely, requests from Columbia were honored marginally more often ($p = .077$ in the one-tailed test of proportions), although there was no significant difference in response rates. Overall, we observed some of the predicted effects, but they seem to be quite weak.

What regards response timing, perhaps the best way to represent it is by means of Kaplan-Meier failure functions, whereby “failure” is a misnomer typical for survival analysis jargon as it actually represents a reply to our request. Figure 1 shows response times, and Fig. 2 shows compliance times.

Table 1: Response and compliance rates by treatment.

Sent from		Not sent by faculty	Sent by faculty
Warsaw	Responded	56%	66%
	Complied	40%	38%
Columbia	Responded	60%	74%
	Complied	44%	54%

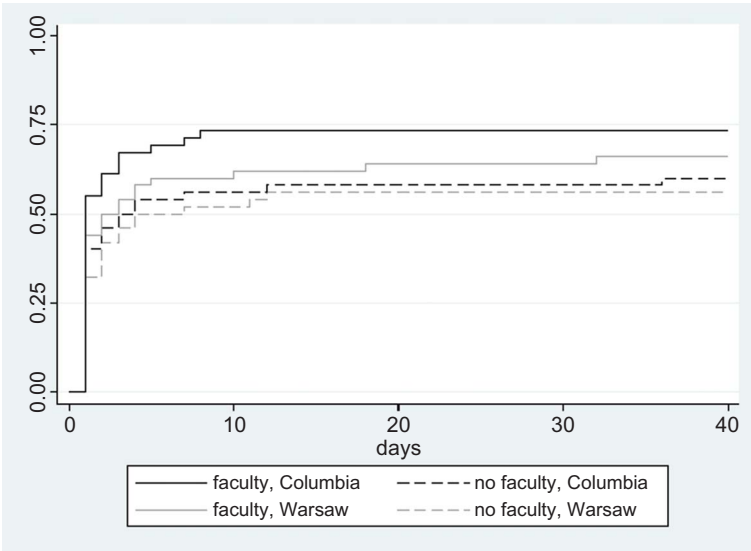


Figure 1: Failure functions for response, by treatment (color figure available online).

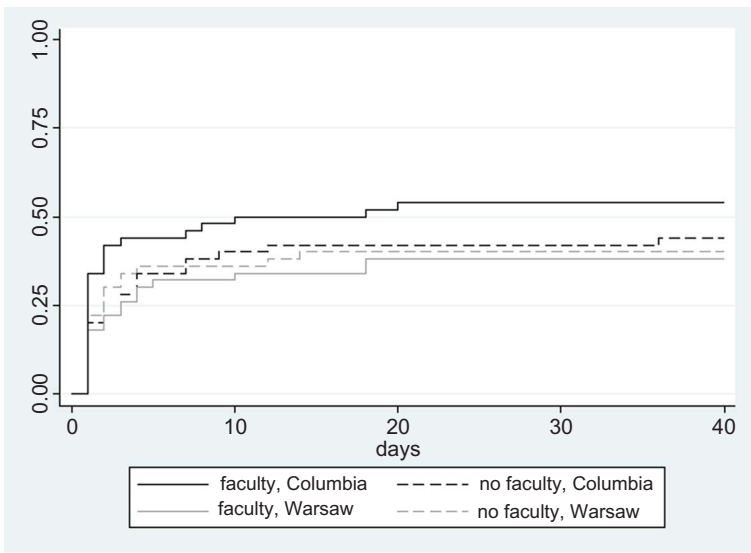


Figure 2: Failure functions for compliance, by treatment (color figure available online).

Again, it seems that being a faculty member, preferably at Columbia University, tends to shorten waiting times slightly. Results of econometric analysis (Cox proportional hazard model) are shown in Tables 2 and 3. Two alternative models are shown: Model 1 only uses treatment variables and the

Table 2: Cox proportional hazard model for responses.

	Model 1	Model 2
Faculty sender	1.294 (.082)	1.339 (.53)
Warsaw	.758 (.175)	.782 (.236)
European author	.944 (.778)	.949 (.798)
Warsaw*European	1.414 (.24)	1.408 (.242)
Female author		0.724 (.105)
Published		1.415 (.046)
Scholar page		1.001 (.502)
Impact factor		.970 (.605)

Note. Hazard ratios reported and *p*-values in parentheses. Impact factor zero assumed for working papers.

Table 3: Cox proportional hazard model for compliance.

	Model 1	Model 2
Faculty sender	1.189 (.385)	1.368 (.12)
Warsaw	.727 (.273)	.780 (.397)
European author	1.409 (.188)	1.529 (.104)
Warsaw*European	1.135 (.747)	1.104 (.913)
Female author		1.100 (.694)
Published		2.279 (.001)
Scholar page		.999 (.833)
Impact factor		.912 (.357)

Note. Hazard ratios reported and *p*-values in parentheses. Impact factor zero assumed for working papers.

postulated interaction between the affiliation of the sender and the author. Model 2 also includes other paper-specific characteristics.

Survival analysis thus suggests possibility of a weak impact of sender's faculty status on the speed of response (note that hazard ratios above one mean positive impact and below one—negative). No significant interaction effect that we hypothesized (positive impact of broad geographic congruence of sender's and recipient's affiliation) is found. As for paper-specific covariates, authors of published papers are much more likely to respond and send the requested material. As a matter of fact, we find that only 32% of requests pertaining to working papers are honored, whereas the fraction is 53% for publications. This is the single most important determinant of our subjects' behavior. Some explicit refusals that we obtained suggest that working papers' authors were reluctant to share materials before publication, possibly to secure credit for being the first to publish a paper incorporating given idea, data set, etc. As put succinctly and openly by one author, “[. . .] so for the short term—I think our description about data access is inac[c]urate—but will be accurate when we get the pub[lication]”.⁴ It is also possible that they do not want to help others find possible weak points of their work before it is accepted or that they consider their own promises made in a working paper not binding.

Other reasons to refuse generally pertained to unavailability of the materials—either due to computer crash, death of the responsible co-author, etc., or confidentiality issues. Overall, however, we only received about 15 explicit rejections. Else, the authors who cared to reply but not to share would revert to rather discouraging utterances such as “I’m traveling [. . .] I should be in a better position to address your query in a couple of months.”

Further analysis showed that categorical variables reflecting the type of object and the field (as proxied by JEL codes) showed overall little impact, although we are maliciously happy to announce that authors of published finance papers complied less often than others for which JEL codes were available (24% vs. 56%, $p = .023$ in a two-tailed test of proportions).⁵

5. CONCLUSION

Overall, the results of our experiment are somewhat reassuring. First, most of the authors responded to our request, and some 44% did fully comply with it. These numbers are higher than in most previous studies, although we did not send reminders. (It may also be that our lack of publications related to a typical paper in our sample decreased response rates, although the threat of finding a mistake or “stealing” a contribution could thereby be reduced). Relatively high response and compliance rates may plausibly have to do with the fact that our sample only included papers in which the promise to share was made explicitly. The authors thus probably felt more responsible to deliver than in the previous studies where the obligation only followed from the journal policy. Those authors thus could justify noncompliance by claiming they had been unaware of the policy (who reads “terms and conditions,” etc., before ticking off anyway?) or felt “forced” to pledge allegiance, for it was a necessary condition to have the paper accepted.

However, we see that authors still take the promises they made themselves quite lightly as long as the paper is not published yet. This calls for a general discussion of the status of working papers. Our findings suggest that authors often take little responsibility for what they write there. For example, a few of the authors said they would only comply at some point in the future (typically not specified precisely), because they are working on a new version. We propose that this is not a valid excuse why one fails to deliver on the promise made in the previous version that she or he chose to make publicly available. Another way of interpreting this tendency is to suppose that some working papers may be of inferior quality and authors are afraid of verification (Wichert et al., 2011). Finally, one can speculate that the admittedly mild pressure that economic journals impose on their authors does make a difference in the sense of increasing the frequency of sharing.

In any case, it is worth emphasizing that authors that responded at all, either complied or tried to find some excuse. Apparently, they did recognize

sharing as a norm, although in some cases sought to avoid its inconvenient consequences.

We were also positively surprised that the impact of affiliation and position of the requestor seemed so weak. These findings suggest that, as far as obtaining access to supplementary materials is concerned, the profession offers a nearly level playing field. It must be noted though, that our test of the impact of the academic position was not very stringent, in that our responders could easily check that we were in fact faculty members (and we cannot verify how many of them did), possibly weakening the treatment effect. An experiment with two different persons (e.g., a professor and a graduate student from the same institution) is of course an attractive alternative solution.

An important novelty of our study was that it covered a broader range of supplementary materials, not just data sets. Interestingly, no difference was found, although the sample was fairly small for all but one category (econometric results), so further research is required.

One could conclude from our study that authors should be encouraged (but not forced) to make explicit statement about availability of data or whatever other materials. Another “soft” measure that may help was suggested by Fischer and Zigmond (2010), advocating explicit credit that authors may receive from their institutions for sharing (when the results are published by somebody else).

However, as others argued before, most effort that needs to be exerted to make supplementary material available (such as writing a codebook) should be taking place as part of the research project anyway. They should thus typically just be made available right away, not sent on request. Our results seem to support the notion that obligatory archiving, enforced by editors as a condition for acceptance, is the only cure against procrastination, opportunism, computer crashes, and other reasons for which resources are now wasted on duplicating already existing files.

NOTES

1. Note that we did not claim in our e-mails that we were working on a related project and would actually consider, e.g., re-use of the obtained stimuli in a subsequent study, although we acknowledge this was likely to be our subjects’ understanding.
2. For the same reason we did not send reminders and we decided to limit ourselves to the sample size of 200, although it would have been easy to expand it if necessary.
3. Partial compliance was rare. Qualitatively, results do not change when these cases are included as positive observations. Occasionally, even more enthusiastic response than full compliance was observed: one author sent the requested material immediately and then, presumably having forgotten about it, he sent it again after three months. Several authors got interested in our projects and asked us to share our findings; one of us was invited to lunch and subsequently to a presentation at the author’s university, suggesting that sending random e-mails may be an effective way of networking.

4. To protect identity of the authors and secrecy of correspondence we cannot offer access to all the e-mails we have received. However, we are willing to prepare a version involving no names, places, obvious references to the paper in question or typical figures of speech possibly enabling recognition of personal style. As this requires additional effort which appears useless from the viewpoint of our study itself, it will be, as the reader may guess, performed upon request.

5. However, testing for all JEL codes, for which we have reasonably large number of observations (when verifying equality of compliance rates we split between two categories that we find to be different—published and unpublished papers) we end up with ten different tests, so a conservative Bonferroni correction would question significance of this result.

ACKNOWLEDGMENTS

The authors are grateful for the efforts taken by the subjects in the course of the study. The first author acknowledges financial support from the Marie-Curie European Reintegration Grant 239172.

REFERENCES

- Dewald, W.G., Thursby, J.G., and Anderson, R.G. (1986). Replication in empirical economics: The journal of money, credit and banking project. *The American Economic Review* 76(4): 587–603.
- Feigenbaum, S. and Levy, D.M. (1993). The market for (ir) reproducible econometrics. *Accountability in Research* 3(1): 25–43.
- Fischer, B.A. and Zigmond, M.J. (2010). The essential nature of sharing in science. *Science and Engineering Ethics* 16(4): 783–799.
- Freese, J. (2007). Replication standards for quantitative social science. *Sociological Methods and Research* 36 (2):153–172.
- Glänzel, W. and Schubert, A. (2005). Analysing scientific networks through co-authorship. In *Handbook of Quantitative Science and Technology Research*. H.F. Moed, W. Glänzel, and U. Schmoch, eds. Dordrecht: Kluwer.
- Kirwan, J.R. (1997). Making original data from clinical studies available for alternative analysis. *The Journal of Rheumatology* 24(5): 822.
- McCullough, B.D. and Vinod, H.D. (2003). Verifying the solution from a non-linear solver: A case study. *The American Economic Review* 93(3): 873–892.
- McCullough, B.D., McGeary, K.A., and Harrison, T.D. (2008). Do economics journal archives promote replicable research? *Canadian Journal of Economics/Revue canadienne d'économie* 41(4): 1406–1420.
- McCullough, B.D., Harrison, T.D., and McGeary, K.A. (2006). Lessons from the JMCB archive. *Journal of Money, Credit, and Banking* 38(4): 1093–1107.
- Savage, C.J. and Vickers, A.J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PLoS One* 4(9): 70–78.
- Smith, E. (2011). The limits of sharing: An ethical analysis of the arguments for and against the sharing of databases and material banks. *Accountability in Research* 18(6): 357–381.

- Trimble, V. and Ceja, J.A. (2011). Are American astrophysics papers accepted more quickly than others? *Scientometrics* 89(1): 281–289.
- Wicherts, J.M., Borsboom, D., Kats, J., and Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist* 61(7): 726.
- Wicherts, J.M., Bakker, M., and Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS One* 6(11): e26828.