Testing the Bounds of the Knowledge Burden Hypothesis: The Fall of the Iron Curtain and Collaboration in Mathematics

Sourobh Ghosh, Mahdi Hashemian, Casey Kearney*

April 27, 2016

Abstract

Prior literature explains increasing team-sizes in academic collaboration as a result of the knowledge burden hypothesis, which broadly asserts that the continuously accumulating stock of scientific knowledge requires increased specialization in academic work. In this paper, we replicate Agrawal et al. (2016)'s analysis of the differential increases in team-sizes of Soviet-rich and Soviet-poor subfields of theoretical mathematics after the fall of the Iron Curtain. We question the underlying parallel trends assumption of existing analysis and examine non-linear models to potentially explain the observed phenomena. In addition, we integrate previously omitted factors into the analysis such as subfield popularity. Overall, our analysis invites potential for factors other than the knowledge burden hypothesis to explain differential increases in team-size in theoretical mathematics. We conclude with discussion of implications of our findings.

Keywords: knowledge burden; economics of science; innovation

*Ghosh: Doctoral Student, Harvard Business School, sourobh@hbs.edu; Hashemian: Doctoral Candidate, MIT Sloan School of Management, mhashemi@mit.edu, Kearney: Doctoral Candidate, Harvard University, ckearney@g.harvard.edu. We gratefully acknowledge the support and guidance of the Gov 2001 teaching staff: Gary King, Mayya Komisarchik, and Anton Strezhnev. Also, thanks to Andrew Leber and Gabrielle Malina on comments on an earlier draft of this work, and to colleagues in our respective programs for their feedback.

1 Introduction

An influential finding to have emerged from the economics of science literature is the existence of a knowledge burden mechanism that motivates the behavior of innovators (Jones, 2009). Given the continuously accumulating stock of total knowledge, it becomes increasingly difficult for successive generations of innovators to acquire and potentially contribute to the existing stock of knowledge. This hypothesis is used to explain a variety of observed phenomena, such as increased age of first invention, the increasing specialization of labor, and increasing trends in academic collaboration, to name a few examples.

In this paper, we seek to test the robustness of the explanatory power of the knowledge burden hypothesis. Recent studies have identified an interesting exogenous shock with which to test this hypothesis: the fall of the Iron Curtain (Borjas and Doran, 2012; Agrawal et al., 2016). In the fields of theoretical mathematics, Agrawal et al. (2016) (hereafter AGT) argue that the dissolution of the Soviet Union caused an outward shift in the knowledge frontier given the sudden access to ideas and knowledge that for various reasons (political, logistical, etc.) remained behind the Iron Curtain and away from consumption of the mathematics community outside the Soviet Union. This outward shift in the knowledge frontier is alleged to be especially salient in subfields in which Soviet authors were relatively research active (i.e Soviet-rich subfields), explaining a disproportionate increase in team-size in publications in Soviet-rich subfields relative to Soviet-poor subfields after the fall of the Iron Curtain. However, several potential alternative explanations for the observed behavior remain, such as the diffusion and adoption of information technology (IT) as tools for collaboration in theoretical mathematics, the relative popularity of certain sub-fields of study, and other potentially unobserved or omitted factors.

We begin our analysis in Section 2 by replicating AGT's difference-in-differences (DID) estimation of increases in co-authorship in theoretical mathematics publications. Using our replication as a baseline, in Section 3 we conduct a variety of tests aimed at examining the robustness of the knowledge burden hypothesis as it is applied to interpreting the observed phenomenon. Overall, we find some support for a potential violation of the non-linear trends assumption to AGT's DID estimation. We also control for subfield popularity as a potential covariate and find AGT's original specification to be robust to this additional control. We then discuss potential implications of our analysis in Section 4.

2 Replicating Agrawal, Goldfarb, and Teodoridis (2016)

We begin by replicating the main results of AGT (Agrawal et al., 2016). Please refer to the original paper for complete details of the authors theoretical motivation, empirical strategy, and other details. Here we shall highlight aspects of the authors original analysis that is pertinent to the extension analysis we present in this paper. Given the exogenous shock of the fall of the Iron Curtain, AGT hypothesize that an observed disproportionate increase in team-size in Soviet-rich subfields of theoretical mathematics can be explained due to an advance in the knowledge frontier in those Soviet-rich subfields. Using methodology established in (Borjas and Doran, 2012), AGT rank Soviet-rich subfields by the proportion of Soviet publications to American publications in a particular subfield of mathematics over the 1984-1989 time frame (thus a Soviet-rich subfield has relatively more Soviet to American publications than a Soviet-poor subfield).

To test their hypothesis, AGT employ a DID strategy between Soviet-rich and Soviet-poor subfields, before and after the fall of the Iron Curtain (defined by the authors as 1990). Data is collected on every theoretical mathematics publication published between 1970 and 2010. To help isolate the effect of the advance of the knowledge frontier in Soviet-rich fields, all publications with at least one Soviet co-author is dropped from this data set (to avoid potential confounding effects). Team-size is operationalized as the log of the author count for a given publication.

We have successfully replicated AGT's results here, presented in Table 1. The first column compares differences in the top three and bottom three fields, whereas the second column compares differences in the top three subfields relative to the remaining subfields. Columns three and four increase the number of subfields considered in the SovietRich variable, and finally the fifth column considers a "continuous specification where the SovietRich variable is represented by the (Borjas and Doran, 2012) ranking of each subfield. While replicating these results, we encountered slight differences in the observation count (difference of one for column one, difference of four in columns two to five). This did not yield materially different results on any quantity of interest as reported in Table 1 relative to AGTs original work.

Table 1 consistently demonstrates evidence for a disproportionate increase in publication team-size for a variety of operationalizations of the Soviet-richness of a particular subfield (except in the continuous case). Recognizing a potential confounding factor in the emigration of Soviet scholars to the US and Europe in particular, AGT examine a geographical context with relatively limited Soviet immigration, namely in Japan. Restricting the data set to publications only in Japanese journals, we successfully replicate AGTs analysis here in Table

Table 1: Teams in Soviet-Rich Subfields Exhibit a Disproportionate Increase in Size After 1990

	Dependent variable: log of author count				
	Top and bottom three subfields	Soviet-rich defined as top three of subfields	Soviet-rich defined as top five of subfields	Soviet-rich defined as top ten of subfields	Continuous
AfterIronCurtain	0.0780***	0.0489***	0.0394***	0.0389**	0.0011
× SovietRich	(0.0117)	(0.0106)	(0.0138)	(0.0152)	(0.0011)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Subfield fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.113	0.106	0.106	0.106	0.106
Observations†	133,498	563,466	563,466	563,466	563,466

Notes: The unit of analysis is the publication. All models are OLS with robust standard errors, clustered by subfield. *p<0.1; **p<0.05; ***p<0.01. †Observation count does not replicate perfectly with Table 3 in paper.

Table 2: Teams in Soviet-Rich Subfields Exhibit a Disproportionate Increase in Size After 1990 (*Top and bottom three subfields from Japanese journals*)

	Dependent variable: log of author count		
	All Japanese journals	Ranked Japanese journals	Unranked Japanese journals
AfterIronCurtain × SovietRich	0.0692***	0.0678***	0.0956**
	(0.0173)	(0.0203)	(0.0206)
Year fixed effects	Yes	Yes	Yes
Subfield fixed effects	Yes	Yes	Yes
R^2	0.076	0.093	0.101
Observations	5,096	3,859	1,237

Notes: The unit of analysis is the publication. All models are OLS with robust standard errors, clustered by subfield. *p<0.1; **p<0.05; ***p<0.01.

2, demonstrating a similar disproportionate increase in team-sizes in Soviet-rich subfields.

To demonstrate further evidence of their proposed hypothesis, AGT examine citation behavior to Soviet prior-art. Please refer to AGT for details of the empirical strategy. We successfully replicate their results and present them in Table 6 in the Appendix.

Overall, AGT present a series of analyses that seem to corroborate their hypothesis that a disproportionate increase in academic collaboration in Soviet-rich subfields is observed due to an outward shift in the knowledge frontier in those areas. However, temporally the fall of the Iron Curtain coincided with a variety of other historical trends, such as the adoption and diffusion of IT and the Internet. While AGT concede the possibility that the differential adoption of IT across subfields may potentially confound their analysis, they largely maintain that some pre-trends would be noticeable prior to the fall of the Soviet Union. However, it is not immediately apparent that theoretical mathematics as a discipline followed the same diffusion and adoption patterns of IT that the engineering disciplines studied in prior work did to support the claim that communication costs were falling in the 1980s (Agrawal

and Goldfarb, 2008). Furthermore, beyond the influence of IT, there may be other factors involved such as the underlying popularity of particular sub-fields, and a host of unobserved factors. The remainder of this paper is dedicated to testing the robustness of AGT's claim that the knowledge burden hypothesis is driving the observed effects of increased team-sizes in Soviet rich subfields of theoretical mathematics.

3 Testing Robustness of the Knowledge Burden Hypothesis

Given the relative parsimony of AGT's estimation strategy, we test the robustness of the implicit claim that the operationalization of $AfterIronCurtain \times SovietRich$ with subfield and year fixed effects (i.e. the models examined in Table 1) solely captures the knowledge burden effect in explaining observed variation in team-size. We begin by testing the robustness of the cutoff year used in the $AfterIronCurtain \times SovietRich$ interaction (beyond the tests employed by AGT). We then present models for the differential exponential increase in team-size over time as a possible alternate explanation of the observed trends. Finally, we incorporate a subfield "popularity" effect into AGT's original estimation strategy to test robustness of influence of the outward shift in knowledge frontier due to the fall of the Iron Curtain.

3.1 Testing the Robustness of the Iron Curtain Fall

In their paper, AGT suggest that having different cut-off points related to the year of the Iron Curtain fall in the interaction term will produce the same result. We tested this and found that replicating the analysis using 1989, 1990 and 1991 will not qualitatively change the results. In addition, AGT suggest they could not find any difference between the two subfield groups (Soviet-rich vs. Soviet-poor) prior to the fall of the Iron Curtain fall, and therefore the collapse of the Iron Curtain is a primary contributor to the change in the number of authors. We extend the DID analysis through a series of placebo tests to a broader set of cut-off points for the year in the interaction term.

In their primary DID test, AGT assign the "cut-off" year for the AfterIronCurtain variable to be 1990, thus representing the fall of the Iron Curtain with a dummy variable that takes the value of 1 after 1990 in the interaction of $AfterIronCurtain \times SovietRich$. Here, we alter the cut-off year to test the robustness of the claim that the fall of Iron Curtain drove differential impacts on co-authorship in mathematics.

We present results from this analysis in Table 3. In the first column, we utilize the

full data set for top and bottom three subfields (corresponding to the model in column 1 of Table 1), altering the cut-off year as 1980 and 2000, respectively. Building on these results, we then drop observations after 1990 in the After1980 specification (cut-off year as 1980) while running the same DID test and drop observations before 1990 in the After2000 specification. Taken collectively, these results would seem to suggest that while there may not be differential trends between Soviet-rich and Soviet-poor subfields in team-size prior to the fall of the Iron Curtain (as AGT argue), we witness significant effects in the post-1990 subset of observations when defining 2000 as a cut-off year. This may potentially call into question the influence of the fall of the Iron Curtain, since if we are to accept this hypothesis, we would expect most of the outward shift in knowledge frontier to have acted prior to 2000, yielding a non-significant result in the $After2000 \times SovietRich$.

One way to interpret these results is that Soviet-rich and Soviet-poor fields have different underlying trends in publication co-authorship. While a DID approach attempts to account for this, it assumes linearity in pre- and post- trends. Given the four decade time-span of the dataset, which potentially includes a number of unobserved or omitted confounders, we consider what the impact would be if the underlying trends were in fact non-linear. In the following section we simulate exponential data generating processes to potentially explain differential co-authorship across Soviet-rich and Soviet-poor fields of mathematics.

Table 3: Testing Different Cut-off Points (*Top and bottom three subfields*)

	Dependent variable: log of author count		
	Full Dataset	Dropping After 1990	Dropping Before 1990
After1980 × SovietRich	0.05817***	-0.0049	
	(0.0173)	(0.00138)	
After2000 × SovietRich	0.09375***		0.07657***
	(0.00869)		(0.0807)
Year fixed effects	Yes	Yes	Yes
Subfield fixed effects	Yes	Yes	Yes

Notes: The unit of analysis is the publication. All models are OLS with robust standard errors clustered by subfield. *p<0.1; **p<0.05; ***p<0.01.

3.2 Alternate Specifications: An Exponential Model

Given the results from Section 3.1, in this section, we focus on possible inherent differences in the Soviet-rich and Soviet-poor fields. The two subfields could be on different trajectories throughout the entire timeframe (1970 2010), but the trajectories could overlap more in the earlier years in the sample. We conducted a simulation experiment to determine if it is

possible to fail to identify the assumed difference between the two fields prior to 1990, while there is a significant change of behavior after 1990. If author counts in the two fields were each increasing at an exponential rate over time, they may lie close to each other in early periods, while naturally diverging over time. A significant difference could then be identified merely be associated with to the passage of time, and not be the result of an exogenous event like the collapse of the Iron Curtain. We modeled this exponential form and further assumed that author counts in the two subfields are explained by a different exponential function but start from the same point at 1970. To complete the simulation, we added normally distributed random noise to the exponential form to simulate the variation in the data. After examining the predicted values and the residuals from AGTs initial specification, there is evidence of an increase in variation of the residuals over time, especially after 1990 (see Figure1). Further, the increase in variation appears stronger in the Soviet-poor subfield[1]. To account for this change in variation, we assigned different levels of variation to the random noise assigned to each of the two types. to see under what circumstances if at all this variation can shadow the inherent change in the exponential functions.

Figure 1: Residuals of Log Author Count by Year

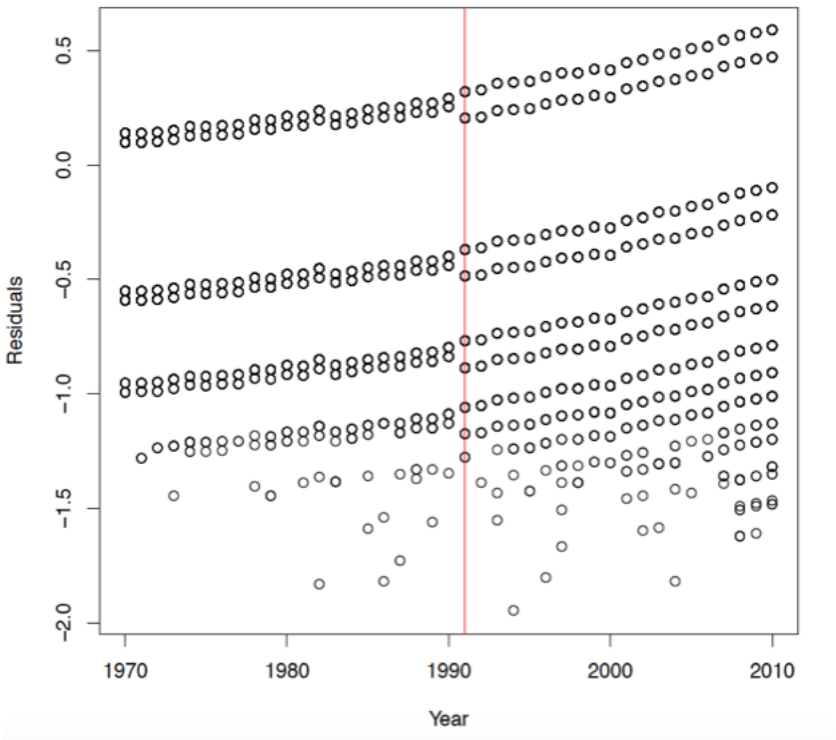

Table 4 presents the results of this analysis. Each model has two components. In the first part we test if results are significant with 1990 as the cut-off year. In all cases, the

[1]See Figure 4 in AGT's original paper

results are significant. However, if we truncate the data and exclude post-1990 in some of the cases (especially when the difference between standard deviations for each type is larger) the regression cannot distinguish between the two types. The simulation attempts to generate data and complete the analysis as close to the real data as possible. For example, we simulated different subfields for each type of field (Soviet-rich vs. Soviet-poor) along with the other necessary data and we calculate robust standard errors in the same manner as AGT employ. In summary, we find that a lack of significant results prior to 1990 does not rule out the possibility of an inherent difference in the underlying determinants of the authorship counts across the subfields. This type of generation process would undermine the authors claim, and additional explanatory variables to consider these effects could help isolate the proposed effect of the DID approach. In other words, a lack of difference prior to 1990 could be an artifact of differences in standard deviations and breach of normality assumption. This concern is borne out empirically in the data, as there is a wide gap between the standard deviation prior to 1990. The collaboration in Soviet-rich fields has a standard deviation of 0.58, while the same measure for Soviet-poor fields is 0.51. This magnitude of difference in standard deviation is in line with the differences used in the simulations.

3.3 Modeling the Effect of Subfield Popularity

In considering alternate explanations for the observed phenomena, we examine the possibility that Soviet-rich subfields may have simply been more popular or robust areas of inquiry than Soviet-poor subfields in the data set of non-Soviet co-authored works, thus obscuring AGT's estimation strategy. Research from the social psychology literature demonstrates a general human preference for social homogeneity (Gruenfield and Tidens, 2010). In the context of academic communities, a desire for cognitive symmetry (among other factors) may draw academics to subfields and areas that are "hot" or popular at the time. This view may also be supported by views of the progress of scientific knowledge, whereby the phase of normal science incentivizes academics to contribute to existing paradigms of theory (Kuhn, 2012).

To consider this case, we construct a new variable to estimate the popularity of a given subject. This popularity measure is defined as the year-over-year increase in the number of publications for a given subfield, or the two year compound annual growth rate in the publication number for the field. Importantly, this measure is separate from what the authors define as a "top" subfield. AGT's definition refers to fields in which Soviet mathematicians have relatively more expertise than their American counterparts. Of course, these top Soviet subfields do not necessarily have to be the most popular areas for publication.

Table 4: Exponential Model Analysis

	Dependent variable: log of author count											
	(1.1)	(1.2)	(2.1)	(2.2)	(3.1)	(3.2)	(4.1)	(4.2)	(5.1)	(5.2)	(6.1)	(6.2)
	SD = [0.62,0.58]	SD = [0.62,0.58]	SD = [0.56,0.58]	SD = [0.56,0.58]	SD = [0.53,0.58]	SD = [0.53,0.58]	SD = [0.51,0.58]	SD = [0.51,0.58]	SD = [0.48,0.58]	SD = [0.48,0.58]	SD = [0.40,0.58]	SD = [0.40,0.58]
After1990Type	0.051**		0.039**		0.036**		0.031**		0.024**		0.019**	
	(0.004)		(0.002)		(0.001)		(0.006)		(0.002)		(0.002)	
After1980Type		0.033**		0.019*		0.013**		0.014**		0.010		0.003
		(0.011)		(0.007)		(0.004)		(0.003)		(0.006)		(0.004)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Journal FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations†	99,893	52,393	99,933	52,434	99,944	52,444	99,944	52,444	99,950	52,451	99,962	52,462

Notes: All models are OLS with robust standard errors clustered by subfield. *p<0.1; **p<0.05; ***p<0.01. The SD is the simulated standard deviation for coauthors of [type1, type2]. Models x.1 are from 1970 to 2010 (complete range), Models x.2 are from 1970 to 1990 (before Iron Curtain fall).

We re-ran the authors main model after including this new variable. This did require us to drop data for the beginning years 1970 and 1971 of the data set, but the baseline specifications without these years match the authors original results. If the popularity metric absorbs the predicted variation from the authors interaction term, it would be an indication that the collapse of the iron curtain had only an indirect effect of co-authorship rates by raising interest levels in this subfield. While this would not rule out the collapse of the Iron Curtain as an important event, it would have important consequences for the theory the authors wish to test. Table 5 below summarizes the results of our analysis.

Model 1 replicates the authors original analysis after dropping all publications from 1970 to allow for the construction of a growth rate variable. The coefficient on the interaction term for this model matches the authors original results in level of significance and magnitude. Model 2 incorporates the year-over-year growth rate of the publications for a subfield into the regression. While this new term is found to be significant (at the 10 percent level), inclusion of the new variable did not alter the coefficient or significance level of the primary interaction term.

Models 3 and 4 consider the two-year compound annual growth rate as the appropriate measure for subfield popularity. A two-year growth rate may be more appropriate if there is a lag time between an increase in popularity and the subsequent publications. Constructing a two-year growth rate requires an additional year of data to be dropped from the beginning of the data set.

Model 3 again replicates the authors original analysis with the earliest two years removed. As before, the coefficient and significance level of the interaction term are not materially changes by this reduction in the sample size. Model 4 incorporates the two-year compound growth rate as an additional explanatory variable. The inclusion of this variable does not have a significant impact on the magnitude or significance of the interaction term. The two-year growth rate is however found to be more significant than the simple year-over-year growth rate tested in Model 2. Together, these tests provide an additional robustness check to the authors finding and show that any potential increase in subfield popularity after the collapse of the Iron Curtain does not fully explain the rise in co-authorship levels in Soviet-rich subfields.

Table 5: Robustness of AGT's Results to the Impact of Popularity on Co-Authorship Rates

	Dependent variable: log of author count for top and bottom three subfields			
	Model 1	Model 2	Model 3	Model 4
AfterIronCurtain	0.0767***	0.0767***	0.0759***	0.0760***
× SovietRich	(0.0112)	(0.0113)	(0.0111)	(0.0111)
Subfield Publication Growth Rate from Prior Year		0.0280* (0.0129)		
Subfield Publication Compound Growth Rate over Previous Two Years				0.1280*** (0.0303)
Year fixed effects	Yes	Yes	Yes	Yes
Subfield fixed effects	Yes	Yes	Yes	Yes
R^2	0.110	0.110	0.107	0.107
Observations	131,938	131,938	130,365	130,365

Notes: The unit of analysis is the publication. All models are OLS with robust standard errors, clustered by subfield. *p<0.1; **p<0.05; ***p<0.01.

4 Discussion and Conclusion

Throughout statistical research, it is important to question the applicability of both the underlying assumptions of the model being utilized and the appropriate quantity of interest one wishes to measure directly or via proxy. The knowledge burden hypothesis highlights an important case where the definition of the quantity of interest is important. AGT assert the number of co-authors is an appropriate variable to measure the degree of required specialization for a successful publication in theoretical mathematics. If this variable is not an appropriate measure of the need for specialization, the significance found for the interaction term relating to the collapse of the iron curtain has little to say to on the actual knowledge frontier hypothesis. To better ascertain this relationship, our analysis considered whether co-authorship rates could be driven by the general popularity of a subfield. If many mathematicians began to focus on newly discovered Soviet-rich subfields after the collapse of the Iron Curtain, the influx of potential co-authors working in the field could be driving the increase in co-authorship independent of any knowledge burden requiring specializations. We find that a popularity measure does provide additional significant predictive value, but fails to lessen the impact of the Iron Curtain interaction term, providing an additional robustness check to the authors conclusions.

Turning to the DID approach employed by AGT, one of the primary assumptions for

this model is that of parallel trends, namely the average change for the non-treatment group (Soviet-poor subfields in this case) is the same as the counterfactual for the treated group. The authors presentation suggests there may be some differential variances between the treated and control groups after the treatment occurs. To test this in the context of the parallel trends assumption, we consider an exponential model that would allow for a change in the variances between the two groups to occur in a similar time frame as the collapse of the iron curtain. This change in the underlying variance could diminish the validity of the parallel trends assumption and question the appropriateness of the DID approach. Finding some evidence to support these concerns, we suggest that future work on this topic should consider alternate estimation strategies, especially given the long time frame of the empirical sample. For instance, our analysis supports the possibility that any time-variant unobserved covariates that roughly coincide with the fall of the Soviet Union (but are otherwise unrelated to this event) may potentially be biasing AGTs estimates.

Of course, our analysis has limitations. The exponential model presented in Section 3.2 represents a first attempt at modeling non-linear trends in co-authorship over time and may not necessarily best represent the underlying data generating process. Furthermore, our operationalization of subfield popularity can be improved in future work (one may, for instance, bring in additional data to assess this beyond AGTs original data set). Limiting our analysis here somewhat was the constraint that our theoretical construction of popularity be unrelated to changes in treatment so as to not invite post-treatment bias. Given that we control for underlying popularity of subfields outside the Soviet Union, changing treatment of Soviet-rich fields should not affect our controls.

Taking the knowledge burden mechanism as a stylized fact has significant policy implications, such as justifying efforts to increase the role of the public sector in incentivizing increased scientific collaboration. We believe it is of crucial importance for scholars to understand the bounds of this mechanism to avoid over-stating its potential influence in explaining phenomena in todays dynamic, fast-changing work environments.

References

A. Agrawal and A. Goldfarb. Restructuring research: Communication costs and the democratization of university innovation. *American Economic Review*, 98(4):1578–90, September 2008. doi: 10.1257/aer.98.4.1578. URL `http://www.aeaweb.org/articles/?doi=10.1257/aer.98.4.1578`.

A. Agrawal, A. Goldfarb, and F. Teodoridis. Understanding the changing structure of sci-

entific inquiry. *American Economic Journal: Applied Economics*, 8(1):100–128, January 2016. doi: 10.1257/app.20140135. URL http://www.aeaweb.org/articles/?doi=10.1257/app.20140135.

G. J. Borjas and K. B. Doran. The collapse of the soviet union and the productivity of american mathematicians*. *The Quarterly Journal of Economics*, 127(3):1143–1203, 2012. doi: 10.1093/qje/qjs015. URL http://qje.oxfordjournals.org/content/127/3/1143.abstract.

D. Gruenfield and L. Tidens. Organizational preferences and their consequences. In *Handbook of Social Psychology*, volume 2, chapter 33, pages 1252–1287. John Wiley and Sons, Inc., 2010.

B. F. Jones. The burden of knowledge and the death of the renaissance man: Is innovation getting harder? *The Review of Economic Studies*, 76(1):283–317, 2009. doi: 10.1111/j.1467-937X.2008.00531.x. URL http://restud.oxfordjournals.org/content/76/1/283.abstract.

T. S. Kuhn. *The structure of scientific revolutions*. University of Chicago Press, 2012.

5 Appendix

Table 6: Teams in Soviet-Rich Subfields Exhibit a Disproportionate Increase in Propensity to Cite Soviet Pre-1990 Art After 1990. The Disproportionate Increase in References to Soviet Prior Art is Driven by Larger-Sized Teams in Soviet-Rich Subfields

	Dependent variable: References to Soviet art			
	Count of Soviet references (defined by Soviet journal	Percentage of Soviet references (defined by Soviet journal	Count of Soviet references (defined by Soviet journal	Percentage of Soviet references (defined by Soviet journal
AfterIronCurtain × SovietRich × LogAuthorCount			0.5628***	0.0167***
			(0.1078)	(0.0029)
AfterIronCurtain × SovietRich	0.4020***	0.0172***	0.1416	0.0097**
	(0.0423)	(0.0036)	(0.0832)	(0.0037)
AfterIronCurtain × LogAuthorCount			0.0558**	0.0033
			(0.197)	(0.0024)
SovietRich × LogAuthorCount			-0.2472**	-0.0122**
			(0.197)	(0.0024)
LogAuthorCount			0.0412**	0.0002**
			(0.0787)	(0.0007)
Year fixed effects	Yes	Yes	Yes	Yes
Subfield fixed effects	Yes	Yes	Yes	Yes
R^2	0.117	0.180	0.128	0.084
Observations	1,012	1,012	1,012	1,012

Notes: The unit of analysis is the publication. All models are OLS with robust standard errors, clustered by subfield. *p<0.1; **p<0.05; ***p<0.01.