Does Knowledge Accumulation Increase the Returns to Collaboration? Evidence from the Collapse of the Soviet Union

Ajay Agrawal, Avi Goldfarb, and Florenta Teodoridis*

October 2012

Abstract

Several theories can explain the increasing size of research teams over time. One of these, the knowledge accumulation hypothesis - that reaching the knowledge frontier is progressively costly over time and results in increasing specialization and collaboration - raises unique policy implications concerning poverty traps, research grant evaluation procedures, and incentives to enter research careers. However, this hypothesis is difficult to separately identify from other explanations for increasing team size that do not raise the same policy implications. We exploit the collapse of the USSR as an instrument because Soviet research was kept hidden from the outside world and then suddenly released. This created a shock to the knowledge frontier due to socio-political events exogenous to the research community. Furthermore, the degree of the shock varies across subfields since Soviet advances were greater in some subfields than others. Using 40 years of publication data in theoretical mathematics covering the period 1970-2010, we employ a difference-in-differences approach to compare the propensity of mathematicians working outside the USSR to collaborate in "Soviet-rich" versus "Soviet-poor" subfields before and after the knowledge frontier shock. Consistent with the theory, we find that team size - the number of coauthors on a paper - increases disproportionately in Soviet-rich subfields after 1990. Furthermore, consistent with the hypothesized mechanism, scholars in Soviet-rich subfields disproportionately increase their citations to Soviet prior art following the shock. Moreover, the knowledge accumulation effect is disproportionately realized through research teams that include a junior scholar, providing insight on the marginal effect of pushing out the frontier on the division of labour.

JEL: O33, O40

---

* Agrawal: University of Toronto and NBER, ajay.agrawal@rotman.utoronto.ca. Goldfarb: University of Toronto, agoldfarb@rotman.utoronto.ca. Teodoridis: University of Toronto, Florenta.Teodoridis09@rotman.utoronto.ca. This research was funded by the Centre for Innovation and Entrepreneurship at the Rotman School of Management, the Martin Prosperity Institute, and the Social Sciences and Humanities Research Council of Canada. We thank seminar participants at the "Organization, Economics, and Policy of Scientific Research" workshop, Carnegie Mellon University, and the University of Toronto for valuable feedback. Errors remain our own.

1. Introduction

Is the accumulation of knowledge increasing the returns to collaboration? Although there is compelling evidence that the returns to collaboration are indeed increasing over time, as evidenced by the steady rise in collaboration, and knowledge is undoubtedly accumulating with the forward march of science, there are several other plausible explanations for the rise in collaboration. It is important to determine whether knowledge accumulation actually increases the returns to collaboration since this explanation raises unique policy implications that are not appropriate in response to the other explanations. Our primary objective in this paper is to present evidence that is consistent with the knowledge accumulation explanation but not with the others. The set of plausible explanations are not mutually exclusive and we do not attempt to rule out the others.

Several prior studies present evidence that team size in research has increased steadily over time (Adams et al, 2005; Wuchty et al, 2007; Jones, 2009). For example, Wuchty et al (2007) show that over the latter half of the twentieth century, team size increased in 170 out of 171 fields in science and engineering, 54 out of 54 fields in the social sciences, and 24 of 27 fields in the arts and humanities. Furthermore, this increase even occurred in fields traditionally associated with individual-oriented research: “Surprisingly, even mathematics, long thought the domain of the loner scientist and least dependent of the hard science on lab scale and capital-intensive equipment, showed a marked increase in the fraction of work done in teams, from 19% to 57%, with mean team size rising from 1.22 to 1.84.” Moreover, they present citation-based evidence that the relative impact of team versus individual output is increasing over time, even after controlling for self-citations.

A number of theories have been advanced to explain this trend: 1) increasing capital intensity, 2) declining communication costs, 3) shifting authorship norms, 4) increasing returns to research portfolio diversification, and 5) an outward shifting knowledge frontier. The capital intensity theory asserts that many fields, such as physics, experienced a rise in capital costs (e.g., particle accelerators), which increase the returns to collaboration due to the indivisibilities of research equipment (Stephan, 2012). The communication cost theory is predicated on the decline in communication and file sharing costs as well as travel costs, increasing the returns to collaboration (Hesse *et al*, 1993; Agrawal and Goldfarb, 2008; Kim *et al*, 2009). The authorship culture explanation is that norms evolved such that contributors who in the past may have been listed in the acknowledgements are increasingly likely to be included as coauthors, especially in lab-based sciences where the lab head is listed last, the lead author first, and other contributors in between (Stephan, 2012). The diversification explanation asserts that as the publication requirements for promotion and tenure increase, there are increasing returns to mitigating publication risk by diversifying one's research portfolio (Stephan, 2012).

The knowledge accumulation theory asserts that successive generations of innovators face an increasing education burden due to the constantly advancing knowledge frontier, which requires innovators to specialize more and thus necessitates working more collaboratively (Jones, 2009). This alters the organization of innovative activity towards more teamwork. Jones provides descriptive statistics consistent with this theory. For example, he shows that over time: 1) the number of co-authors on academic publications increases, 2) Nobel laureates are older when they perform their great achievement, 3) the number of co-inventors per patent increases, 4) the age at first innovation increases, and 5) the probability of switching fields decreases. However, these statistics are also consistent with other explanations, such as increasing capital costs and declining communication costs.

While these explanations need not be mutually exclusive, it is instructive to know whether the outward shifting knowledge frontier does in fact influence the propensity to collaborate since this would raise certain unique policy implications. For example, Jones (2011) presents a model in which the "knowledge burden" leads to a poverty trap. As the knowledge frontier shifts outwards, individuals compensate by specializing and thus the returns to collaborating increase. However, in economies where the market for complementary skills is thin, individuals are less likely to invest in the human capital necessary to reach the frontier. This results in an increasingly thin market for specialized skills (i.e., a trap).

One policy prescription from this theory is to subsidize skills development in a concentrated area (e.g., infectious diseases) in order to address the complementary skills shortage for a finite period of time until the private returns to acquiring specialized skills are sufficient for the labour market to sustain the cycle without further intervention. This policy initiative is not appropriate if knowledge accumulation does not increase the returns to collaboration and the observed rise in team size is actually due to other factors in the economy, such as rising capital costs and/or falling communication costs.

In a separate paper, Jones proposes policies involving changes to the way in which ideas are evaluated (Jones, 2011). If research teams, rather than individuals, are needed to work on scientific problems due to an outward shifted knowledge frontier, then perhaps evaluation teams rather than individuals are needed to evaluate grant applications. Again, this policy prescription is not relevant if the observed increase in team size is not due to knowledge accumulation, but rather to other factors. For example, if team size is increasing due to increasing capital costs, this does not imply increasing returns to team-based evaluation since the equipment is not required for evaluating the grant proposal. Jones also proposes increased subsidies for individuals who enter into careers in science since, under the knowledge accumulation hypothesis, researchers bear increasing private costs to reach the frontier. Again, this policy prescription is not appropriate if the knowledge accumulation hypothesis is incorrect and the observed rise in collaboration is due to other factors.

Therefore, identifying a causal relationship between an outward shift in the knowledge frontier and an increase in the propensity to collaborate, separately from other explanations for increasing collaboration, is important since the knowledge frontier theory has some unique policy implications. However, identification is difficult since many unobservables may be (and likely are) correlated with both collaborative behaviour and time. In order to provide more compelling (albeit not definitive) evidence that an increasing knowledge burden leads to a growing propensity to collaborate, we need an instrument that is correlated with a shift in the knowledge frontier but not with collaboration except indirectly through its effect on the frontier.

The collapse of the Soviet Union in 1989 provides such an instrument. Although the USSR was a world leader in certain areas of science, including various subfields of mathematics, Communist government officials forced their researchers to work in isolation from the rest of the world. For example, with few exceptions scholars were strictly prohibited from traveling, publishing outside of the Soviet Union, and accessing foreign publications without case-by-case government approval. Thus, when the Iron Curtain fell and Soviet science became widely available, the knowledge frontier outside the USSR experienced a shock.

Furthermore, the degree of the knowledge shock across subfields of mathematics varied since Soviet researchers had made significant advances relative to the rest of the world in some subfields but not others. We focus on theoretical mathematics. Borjas and Doran (2012) show that the Soviet mathematics community was very advanced relative to the West in some subfields of theoretical mathematics, such as "partial differential equations" and "operator theory," and much less so in others, such as "abstract harmonic analysis" and "sequences, series, summability."

We exploit this variation in the degree of knowledge shock across subfields using a difference-in-differences type of analysis. Specifically, we compare the propensity of mathematicians working outside the USSR to collaborate in "Soviet-rich" versus "Soviet-poor" subfields before and after the shock. We do this using 40 years of publication data in theoretical mathematics covering the period 1970-2010, 20 years before and after the collapse of the Soviet Union.

We categorize papers using the internationally recognized Mathematics Subject Classification codes developed and assigned by the Mathematical Reviews division of the American Mathematical Society. We follow the Soviet-rich versus Soviet–poor subfield classification developed by Borjas and Doran (2012), which is based on the fraction of publications produced by Soviet researchers during the period 1984-1989. We then focus our attention on mathematicians working outside the USSR and drop observations that involve collaboration with Soviet researchers.

We find that team size - the number of coauthors on a paper - increases after the fall of the Iron Curtain, in both Soviet-rich and -poor subfields. However, consistent with the theory, team size grows disproportionately more in Soviet-rich subfields after the shock. These results are robust to various definitions of Soviet-rich versus -poor subfields. Furthermore, we show that the disproportionate increase in team size in Soviet-rich fields does not begin until shortly after 1990 and then grows over the following 15 years; we find no evidence of a pre-trend. Moreover, we find that authors in Soviet-rich fields disproportionately increase their propensity to draw upon Soviet knowledge after 1990, providing further evidence consistent with the knowledge accumulation explanation.

We then turn our attention to team composition. We report evidence suggesting that junior scholars seem to play an important role in realizing the knowledge frontier effect. The effect is amplified when at least one member of the research team is in the youngest 10 percentile of the age distribution for that subfield-year. We speculate that collaborating with newly trained scholars is an efficient way for research teams to assimilate new knowledge since researchers in training mode can learn new concepts at a lower cost. We also find that the effect is muted when at least one member of the research team is in the oldest 10 percentile of the age distribution for that subfield-year. Perhaps more experienced scholars have greater absorptive capacity and thus are more easily able to assimilate new knowledge themselves and thus receive less benefit from collaboration. Furthermore, consistent with this interpretation, we find that the effect is also muted for teams that include at least one star (a researcher in the 90th percentile with respect to productivity as measured by papers per year). These results begin to shed light on the margins that are most responsive to knowledge accumulation and how this relates to the division of labor.

The remainder of the paper is structure as follows. In Section 2 we provide historical context for our instrument, explaining how knowledge was developed in the Soviet Union and yet kept secret from Western mathematicians, creating the conditions for the 1990 shock to the frontier. In Section 3 we describe our differences-in-differences empirical strategy where we compare the propensity to collaborate in Soviet-rich versus –poor subfields before and after the knowledge shock. In Section 4 we describe the mathematics publication data we use to construct our sample as well as the method we employ for classifying subfields as Soviet-rich or –poor. We present our results in Section 5 and then our conclusions in Section 6.

2. The "Treatment"

Our empirical strategy relies on the assertion that the collapse of the Soviet Union around 1990 caused an outward shift in the knowledge frontier in mathematics and that it did so more for some subfields than others. We rely on three observations to substantiate this assertion: 1) the Soviet Union's effect on the knowledge frontier in mathematics was significant, 2) the Soviet Union's effect on the knowledge frontier was greater in some subfields than others, and 3) the knowledge produced in the Soviet Union was kept secret from the outside world such that its effect on the frontier came reasonably suddenly in the years following 1990. We offer historical context for each of these three points below.

The first observation is that the Soviet Union's contribution to knowledge in the field of mathematics was meaningful and significant. The nation was and continues to be a worldrenowned center of scientific research with mathematics holding a prominent position. Lauren Graham, a historian of Soviet science and technology states: “Of all fields of knowledge, it was mathematics to which Russia and the Soviet Union made the greatest contributions. The Soviet Union became a world power in mathematics. Indeed, Moscow is probably today the city of the greatest concentration of mathematical talent anywhere. The main competitor is no doubt Paris, since mathematicians in the United States, another leader in mathematics in the last generation, are more widely distributed geographically than in France or the Soviet Union.” (Graham, 2008) Graham attributes the nation’s strength in mathematics Scholarly research in mathematics to the fact that it attracted great minds; it was uniquely detached from politics, conferred status and prestige, and offered financial rewards superior to many other occupations.

The second observation is that the Soviet Union’s contribution to knowledge was significantly greater in some subfields of mathematics than others. Borjas and Doran (2012) show this empirically by comparing across subfields the fraction of Soviet to American papers published during the period 1984-1989. We provide further evidence below by comparing the fraction of Soviet to non-Soviet world wide papers published during the period 1970-1989. Graham (1993) notes that although Soviet mathematics was strong across the entire spectrum of theoretical and applied mathematics, they seemed to have made the greatest advancements, relative to the rest of the world, in pure theory.

One explanation for this is politics. Soviet policies were strict about secrecy and focused on maintaining control over technological development. It was easier for Soviet mathematicians to build on their progress in pure theory than in areas where technology implementation was more immediate. Many advances in applied mathematics were stalled for political reasons[1] (Graham, 1993). Differences in subfields were further amplified due to path dependency (Borjas and Doran, 2012); subfields that attracted bright minds early on were more likely to subsequently attract more bright minds due to mentorship opportunities. The importance of mentorship is well known in Science (Merton, 1973) and was likely particularly salient in this setting due to restrictions on travel and access to foreign journals. For example, the success of Moscow mathematics can be traced back to Ergorov and his student N. N. Luzin (Tikhomirov, 2007). Luzin, whose famous work was mainly focused on the theory of functions, mentored subsequent generations of eminent Soviet mathematicians like A. N. Kolmogorov, who "was probably the best-known Soviet researcher on probability" (Graham, 1993). On the other hand, little outstanding mentorship was available to practitioners of some other subfields of theoretical mathematics, like algebraic geometry (Borjas and Doran, 2012).

The third observation is that the knowledge produced in the Soviet Union was kept secret from the outside world such that its effect on the frontier came reasonably suddenly in the years following 1990. Soviet researchers were prevented from publishing their findings, traveling to conferences, communicating or collaborating with non-Soviets, and even accessing non-Soviet references. The Communist government kept very strict control on international travel, regardless of the reason for travel. Academics who wished to attend foreign conferences had to go through a stringent[2] and lengthy approval process and even then most approvals were only granted for travel in Eastern Europe (Ganguli, 2011).

---

[1] Exceptions are directly linked to government interests, such as the space program (Graham, 1993).

[2] The government required many personal details. Those with “tainted” backgrounds were blacklisted for travel (Polyak, 2001).

Obtaining academic publications from outside the USSR was equally painful. Scientists had to place requests with the government for select publications. The Communist party would approve or deny the request based on political criteria that were frequently unrelated to academic considerations, such as suspicions of negative connections with internal political beliefs. Even when an external publication was approved, the government handled the translation process (Polyak, 2001). Attempts to publish or communicate with academics outside the Soviet Union were also subject to similarly harsh scrutiny (Polyak, 2001; Borjas and Doran, 2012). Although these restrictions applied to all countries outside the USSR, the Soviet Union was particularly adamant in severing communication with Western countries. This was particularly isolating since the other leading mathematical research was located in the Western world. Specifically, the US accounted for 51% of total research in mathematics in the 1980s, followed by France with 10% and West Germany with 6.5% (Dubois et al, 2011).

Not only did Soviet knowledge in mathematics take shape with very limited influence from and awareness of related undertakings outside the USSR, but it also remained in relative secrecy until the collapse of the Soviet Union. First, the USSR government kept much of Soviet science secret (Graham and Dezhina, 2008). Second, what escaped the secrecy filter was subject to the natural barrier imposed by the Russian language. Graham and Dezhina (2008) note "the Russian language was known by few researchers outside the Soviet Union, and consequently the achievements of Soviet researchers were more frequently overlooked than those presented in more accessible languages."

Furthermore, mathematics is an area where scholars rely heavily on face-to-face interaction. Walsh and Bayma (1996) report that of all science disciplines, mathematicians need travel the most. Virtual communication is employed only after mathematicians meet in person a few times and have a project well under way. Thus, the travel restrictions on Soviet mathematicians had a strong negative impact on their propensity to collaborate with non-Soviet scholars, further explaining why Soviet mathematics developed in such isolation.

Finally, when the Soviet Union collapsed, Soviet discoveries did begin to spread through the West and were considered new and important. Communication and travel restrictions were lifted, publications were translated and indexed, and ideas and knowledge began to flow out from the Soviet Union into the broader research community. The following quote, from an article published on May 8, 1990, in the *New York Times*, provides an indication of the sudden outward shift of the knowledge frontier.

> *Persi Diaconis, a mathematician at Harvard, said: ''It's been fantastic. You just have a totally fresh set of insights and results.'' Dr. Diaconis said he recently asked Dr. Reshetikhin for help with a problem that had stumped him for 20 years. ''I had asked everyone in America who had any chance of knowing'' how to solve a problem of determining how organized sets become disorganized, Dr. Diaconis said. No one could help. But Dr. Reshetikhin told Dr. Diaconis that Soviet scientists had done a lot of work on such problems. ''It was a whole new world I had access to,'' Dr. Diaconis said.*

In sum, we argue that the fall of the Iron Curtain provides a plausible natural experiment differentially affecting the knowledge frontier across subfields of theoretical mathematics. This historical event was exogenous to the mathematics research community and set free a large pool of accumulated knowledge to contribute to global advances in mathematics. Furthermore, Borjan and Doran (2012) present comprehensive evidence indicating that the timing of the collapse took the global mathematics community by surprise; even in the late 1980s, both the Western mathematical community and USSR scholars were quite certain that Soviet mathematics would remain secluded for the foreseeable future.

3. Estimation Strategy

We employ a difference-in-differences estimation strategy. We use this approach to compare collaboration rates in subfields where the knowledge frontier was most affected by Soviet knowledge ("treated") with subfields least affected ("control"), both before and after the fall of the Iron Curtain. In other words, we can see this as examining the difference between treated and control subfields in two periods, before and after the treatment. This helps us attempt to distinguish between the rise in team size directly attributable to the shift in the knowledge frontier and unobserved differences between treated and control subfields.

The objective of our empirical analysis is to estimate the effect of the knowledge shock on collaboration, which we measure as a count of the number of unique authors on a publication. Thus, we estimate the following linear regression model using the academic paper as our unit of analysis:

$$TeamSize_{it} = \beta(SovietRich_i \times AfterIronCurtain_t) + Subfield_i + \gamma_t + \varepsilon_{it} \qquad (1)$$

$TeamSize_{it}$ is the count of authors for each academic paper *i* published in year t. $SovietRich_i$ is an indicator variable equal to 1 if academic paper *i* belongs to the treated group and 0 otherwise. $AfterIronCurtain_t$ is an indicator variable equal to 1 if academic paper *i* is published after 1990 and 0 otherwise. This applies to academic papers in both treated and control subfields. We include subfield and time fixed effects, hence the main effects $SovietRich_i$ and $AfterIronCurtain_t$ drop out of the estimating equation.

We are primarily interested in the estimated coefficient on the interaction between $SovietRich_i$ and $AfterIronCurtain_t$, which equals 1 for publications in treated subfields that were published after the knowledge shock and equals 0 for all others. We interpret a positive estimated value of this coefficient as implying that the average team size of Soviet-rich subfields increased disproportionately, relative to Soviet-poor subfields, after the knowledge shock, consistent with the knowledge frontier theory.

4. Data

We follow three main steps in the collection and preparation of our data set. First, we extract publication data, then we rank subfields in mathematics with respect to the relative contribution by Soviets, and finally we process the data for analysis. We describe each step in turn.

4.1 Data Collection

We collect data on every publication in theoretical mathematics published during the 40-year period 1970 – 2010. This represents 20 years of data before and after the collapse of the Soviet Union. We follow Borjas and Doran's (2012) interpretation of historical events that isolates 1990 as the year when academic seclusion was significantly lessened. We recognize that the political and social turmoil preceding and following the fall of the Iron Curtain spanned a period of roughly three years, between 1989 and 1991. Our results are robust to choosing 1989 or 1991 as the cutoff rather than 1990.

We collect these data from the American Mathematical Society (AMS). The Mathematical Reviews (MR) division of AMS maintains a comprehensive bibliographic database of worldwide academic publications in mathematics. The MR database includes all mathematics-related journal publications covering the three main categories of mathematics: mathematical foundations (including history and biography), pure or theoretical mathematics, and applied mathematics.[3] Our focus is on theoretical mathematics, which includes analysis, algebra, and geometry (Figure 1).

4.2 Classification

Our empirical strategy relies on exploiting variation in the degree to which the knowledge frontier is shifted outwards as a result of the collapse of the Soviet Union. Specifically, we distinguish between subfields of theoretical mathematics where the Soviets were particularly strong in the years prior to the collapse versus subfields where they were not. We credit Borjas and Doran (2012) for their insight on how to assemble these data to estimate this variation.

We rely on the careful and exhaustive work of the Mathematical Reviews division. They painstakingly classify each paper in mathematics using Mathematics Subject Classification

[3] There are various ways of grouping research areas in mathematics. The one mentioned here follows the MSC code sequencing.

(MSC) codes. The MSC schema is internationally recognized and facilitates targeted searches on research subjects across all subfields of mathematics. The MR team assigns precisely one primary MSC code to each academic publication uploaded to the MR database. There are a total of 40 active primary MSC codes (14 algebra, 19 analysis, 7 geometry) that comprise the theoretical mathematics group.

We drop the six subfields that do not exist throughout the 40-year duration of our study period as well as one subfield for which we are not able to obtain the full data, leaving us with 33 subfields within theoretical mathematics. Next, we adopt Borjas and Doran's (2012) ranking of the remaining 33 subfields, which is based on the degree to which Soviets contributed to a particular subfield. They construct their rank by calculating the ratio of Soviet to American publications in the subfield over the period 1984-1989. (Pre-1990 Soviet publications were added to the MR database after 1990.) They define a publication as Soviet if at least one author has a Soviet institutional affiliation and similarly for American. We list the 33 subfields and their rank in Figure 1.

In addition to using Borjas and Doran's rank, we also construct our own. Ours differs from theirs on three dimensions. First, they define a publication as Soviet based on author affiliation data, whereas we define a publication as Soviet based on author name data.[4] Second, they calculate their rank based on the Soviet to US publication ratio whereas we calculate our rank based on the Soviet to rest of world ratio since we are focused on the overall effect on the knowledge frontier. This seems to make little difference since the US had the

[4] We identify Soviet last names based on conversations with experts and documented rules regarding Soviet surname endings. We then test and calibrate our algorithm by manually looking up and verifying if academics identified as having Soviet last names are indeed Soviets.

largest share of contributions in mathematics, accounting for 51% of total publications between 1970 and 1989 (Dubois et al, 2011). Finally, they determine their rank based on ratios calculated using data six years of data (1984-89) whereas we determine our rank based on ratios calculated using twenty years of data (1970-89). In the end, the ratios are reasonably similar; our ranking classifies the 33 subfields within the same top/middle/bottom groups as theirs. We use the Borjas and Doran (2012) ranking throughout the paper and show that the main results are also robust to our alternate ranking.

4.3 Data Processing

We drop all Soviet publications from the sample. We define Soviet publications as those with at least one Soviet author. We do this to avoid potential confounding effects. After 1990, not only was Soviet knowledge set free to contribute to global advancements in mathematics, but collaboration restrictions were also lifted for Soviet mathematicians. By excluding publications with at least one Soviet author, we account for the possibility of increased co-authorship rates due to removing the constraint previously preventing collaborating with Soviets.

After dropping Soviet publications, our sample includes 563,462 publications spanning the 40-year period. In most of our reported results we focus on a comparison between the three top (Soviet-rich) and bottom (Soviet-poor) ranked subfields, which represent 133,497 publications. However, we show the results are robust to alternative definitions of Soviet-rich: 1) top five ranking subfields, 2) top quartile ranking subfields, and 3) top half ranking subfields and we confirm robustness of all main results to the full sample.

5.0 Results

5.1. Descriptive Statistics

We calculate the mean team size before and after the collapse of the Soviet Union. For the "treated" subfields (Soviet-rich) the mean team size for the 20-year period before 1990 is 1.34 compared to 1.78 for the 20-year period after. By comparison, for the "control" subfields (Soviet-poor) the mean team size is 1.26 before compared to 1.55 after. These differences in means suggest a disproportionate increase in team size for Soviet-rich subfields after the collapse of the Soviet Union (Figure 2).

Specifically, the mean team size for Soviet-rich subfields was just 6% higher than for Soviet-poor before 1990, but 15% higher after. This finding is consistent with the knowledge frontier effect. However, there may be systematic differences between Soviet-rich and -poor subfields that are not accounted for when comparing these simple means. Therefore, we turn to multivariate analysis and our difference-in-differences estimation to study the relationship further.

5.2 Main result: Disproportionate increase in team size in Soviet-rich subfields after 1990

We report the estimated coefficients of Equation 1 in Table 2. We present our main specification in column 1. The key result is the estimated coefficient on the interaction term *AfterIronCurtain**SovietRich*, which is positive and statistically significant. This implies that the difference in average team size between papers in Soviet-rich versus -poor subfields is greater after the shock than before.

We do not present estimates of the main effects of *AfterIronCurtain* or *SovietRich* because these terms are dropped from the estimating equation due to the year and subfield fixed effects. Also, we cluster our standard errors by Soviet rich/poor and before/after 1990. We cluster to address the possibility that shocks experienced in Soviet-rich/poor fields (or before/after) are correlated, for example because, within Soviet-rich/poor subfields, the number of publications in a given year is limited by journal space. As suggested by Donald and Lang (2007) and Bertrand, Duflo, and Mullainathan (2004), failure to take into account such correlations in the errors may lead to an overstatement of the precision of the results, and therefore may inappropriately yield statistically significant results.

Next we show that this main result is robust to various definitions of Soviet-rich in the subsequent three columns. In the second column, we define Soviet-rich as the top five subfields (and Soviet-poor as the bottom five subfields), which expands the number of observations in the sample by 53%, from 133,497 to 204, 470. The point estimate is smaller, but remains positive and significant. In the third and fourth columns, we define Soviet-rich as including all subfields in the top 25 and 50 percentiles, respectively, and Soviet-poor as all subfields not included in Soviet-rich. This expands our sample to the full dataset, increasing our sample size by 322% relative to the original, from 133,497 to 563,462 publications. The main result persists.

However, perhaps team sizes in Soviet-rich subfields increased faster over time than in Soviet-poor subfields for some reason other than the knowledge frontier effect? Although we can think of no plausible explanation for such a difference, if there is one then it could generate our result since the average difference in team sizes in Soviet-rich versus –poor subfields would be larger in the latter 20-year period than the former. To check for such a possibility, we examine the timing of the relationship between the collapse of the Soviet Union and changes in the relative team size in Soviet-rich subfields. Specifically, we run a similar regression to the one shown in Table 2 column 1; however, we replace the single interaction *AfterIronCurtain*SovietRich* with a sequence of dummy variables representing each year interacted with *SovietRich*.

We present the results in Figure 3. Each point represents the coefficient value on the covariate *Year*SovietRich* and thus describes the relative difference in collaboration rates between Soviet-rich and -poor fields in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 1970. We also present these results in table form in Appendix Table 2. The most notable insight from Figure 3 is that the difference between team sizes in Soviet-rich and -poor fields is stable between 1971 and 1990. Then, starting in 1990, the difference in average team size begins to increase, as evidenced by the higher coefficients. Interestingly, the difference in team size seems noticeable after about eight years and then continues to increase for the twelve remaining years in the sample.

In considering this temporal evidence, one might be worried that some factor other than the sudden outward shift in the knowledge frontier might be responsible for generating the estimated coefficient on the interaction. For example, Agrawal and Goldfarb (2008) document the diffusion of Bitnet, an early version of the internet, throughout the 1980s and illustrate how that was related to increased collaboration among researchers. One might worry that the effect of lowered collaboration costs, although spread out over many years and during a slightly earlier period than the 1989-1991 events in the Soviet Union, could explain the result. However, to do so one would have to explain why lowered communication costs systematically affected Soviet-rich subfields more than Soviet-poor. We are not aware of any such explanation.

5.3 Disproportionate increase in Soviet-rich propensity to cite Soviet prior art after 1990

Next we provide further evidence that is consistent with the mechanism underlying the assertion that the collapse of the Soviet Union did in fact generate a knowledge shock. To do so, we turn our attention to citation data. The intuition is that if the lifting of publication restrictions did indeed shift the knowledge frontier outwards, and more so in Soviet-rich fields, then this should be observable through researchers in Soviet-rich subfields disproportionately increasing their propensity to cite Soviet prior art after 1990.

To accomplish this we collect data on references for a subsample of our data. Specifically, we collect backward citation data for papers from the five top and five bottom subfields that are published in one of the top 30 journals of mathematics (as measured by impacts factor and h-index). We further restrict the data to a window of three years before and after the collapse of the Soviet Union (1998-1993) for tractability (this data collection process is manual). We extract 1,214 publications that meet these criteria and are authored by non-Soviet scholars.

Next, we search for these publications in the Web of Knowledge reference database maintained by Thompson Reuters. We find full text information on 917 papers for which we extract the list of references. We count references to Soviet prior art and calculate the percentage of Soviet references relative to the total number of references. We define a citation (prior art) as Soviet if at least one of the authors has a Soviet last name as identified by our name algorithm. We check the robustness of our finding by using an alternative definition where we define a citation as Soviet if it is published in a Soviet journal.

We use these data to estimate a difference-in-differences linear regression, similar to the one estimated in Section 5.2 above, but this time employing a measure of citations to Soviet prior art as the dependent variable:

$$SovietArt_{it} = \beta(SovietRich_i \times AfterIronCurtain_t) + Subfield_i + \gamma_t + \varepsilon_{it}$$

We report our estimated coefficients in Table 3 and our main specification in column 1. In this case we define $SovietArt_{it}$ as a count of the number of references to Soviet citations (defined by name) by academic paper *i* published in year *t*. The estimated coefficient on the interaction *SovietRich*AfterIronCurtain* is positive and significant, implying that researchers publishing in Soviet-rich fields disproportionately increase their propensity to cite Soviet prior art after 1990, relative to those publishing in Soviet-poor fields.

Next, we show that this result is robust to multiple definitions of "Soviet prior art." In column 2 we define the dependent variable as the percentage of total references to papers that have at least one Soviet author identified using our last name algorithm. In column 3 we define the dependent variable as a count of references to papers published in Soviet journals and in column 4 as the percentage of papers published in Soviet journals relative to the total number of references. In all three cases, the main result persists.

5.4 From individual to collaborative research

Our findings could reflect the propensity for researchers that were already likely to be collaborative to simply form larger research teams after the knowledge frontier shift (intensive margin). However, our estimated coefficient may also be picking up a tendency of researchers publishing in Soviet-rich subfields to disproportionately, relative to Soviet-poor, shift from individual to collaborative research after 1990 (extensive margin). To examine this issue we estimate the same differences-in-differences equation as in Section 5.2 above, except use *Collaboration* as the dependent variable. Collaboration is a dummy variable that takes a value of 1 for papers with more than one author and zero otherwise.

We report the main result in Table 4, column 1. The estimated coefficient on the interaction term *AfterIronCurtain***SovietRich*, which is positive and significant, indicates that the likelihood of papers being coauthored (as opposed to single authored) is disproportionately higher after 1990 for those in Soviet-rich relative to -poor subfields. This result is based on our main sample where we define Soviet-rich (-poor) as the three top (bottom) ranked subfields. We check the robustness of this result to alternative definitions of Soviet-rich: 1) top 5 subfields (Table 4, column 2), 2) top 25% of subfields (column 3), and 3) top 50% of subfields (column 4). In each case, the main result persists. Our analysis of whether there is a disproportionate increase in Soviet-rich team size on the intensive margin is inconclusive.

5.5 The Frontier Effect is Disproportionately Realized through Teams with Junior Scholars

We now turn our attention to how the knowledge frontier effect relates to team composition. We begin by examining the age of collaborators. To examine this issue we construct a dummy variable, *Age*, that indicates whether at least one coauthor on a project is in the youngest decile of the age distribution for that subfield-year. To do this we calculate the age of each author on a paper by taking the difference between the year of the focal publication and the year of that author's first publication in the same subfield.

We estimate a triple difference, interacting the *Age* indicator variable with the indicators for Soviet-rich and post-Soviet (*AfterIronCurtain***SovietRich***Age*). We report the result in Table 6 (column 1); it indicates that Soviet-rich teams that include a junior scholar are likely to be disproportionately larger after 1990 relative to Soviet-rich teams that do not include a junior scholar. The result holds when we define *Age* as an indicator that equals one when at least one coauthor on a project is in the youngest *quartile* of the age distribution for that subfield-year (column 2). Finally, we show that the sign of the estimated coefficient on the three way interaction flips when we define Age as an indicator that equals one when at least one coauthor on a project is in the oldest decile (column 3) or quartile (column 4) of the age distribution for that subfield-year.

There are at least two non-mutually exclusive explanations for this result. First, it could be that collaborating with newly trained scholars is an efficient way for research teams to assimilate new knowledge since researchers in training mode can learn new concepts at a lower cost. Or, it could be that more experienced scholars have greater absorptive capacity and thus are more easily able to assimilate new knowledge themselves and thus receive less benefit from collaboration. The latter explanation receives additional report from our next result.

5.6 The Frontier Effect is Disproportionately Weak for Teams with Stars

Given the important role of research stars in the innovation literature (Merton, 1973; Rosen, 1981; Zucker *et al*, 1998; Azoulay *et al,* 2010; Singh and Fleming, 2010; Oettl, 2012; Waldinger 2012) we examine whether teams that include a star are more or less likely to be influenced by the knowledge frontier effect. Following from the argument raised in the prior section, it could be the case that star researchers have high absorptive capacity and are thus able to assimilate new knowledge at low cost and therefore derive less benefit than others from increased collaboration. On the other hand, star researchers may have more resources than others and therefore be more easily able to collaborate (since collaboration is costly) when the returns to doing so increase. So, the direction in which stars influence the frontier effect on team size is an empirical question and thus we take it to the data.

In doing so we construct a dummy variable, *Star*. To do this we calculate the age of each author as in the prior section. We similarly calculate the stock of publications for each author by counting the number of publications in the focal subfield on which that individual is an author between 1970 and the focal year. We calculate each individual's productivity as their stock divided by their age. For example, an individual with a stock of 20 papers that has a research age of 10 years has a productivity value of 2 (20/10). We set the indicator variable, *Star*, equal to one for a paper if at least one author on the paper has a productivity value in the 90th percentile of the productivity distribution for that subfield-year.

We estimate a triple difference, interacting the *Star* indicator variable with the post-Soviet and Soviet-rich variables (*AfterIronCurtain***SovietRich***Star*). We report the result in

Table 7 (column 1); it indicates that Soviet-rich teams that include a star researcher are likely to be disproportionately smaller after 1990 relative to Soviet-rich teams that do not include a star. The result holds when we define *Star* as an indicator that equals one when at least one coauthor on a project is in the 75th percentile of the productivity distribution for that subfield-year (column 2). This result implies that teams with stars experience a lower return to increasing their team size as a result of the outward shift in the knowledge frontier, perhaps since stars are better able to assimilate new knowledge themselves and thus benefit less from further collaboration.

6. Discussion and Conclusion

We find evidence that an outward shift in the knowledge frontier is associated with a subsequent increase in research team size. Importantly, we present evidence that is consistent with the knowledge frontier explanation but not so for other explanations. In other words, although this evidence is not intended to (and does not) rule out other explanations, as they are not mutually exclusive, it does suggest that the knowledge frontier explanation accounts for at least some of the increase in team size that is widely documented.

In the case of mathematics, a back-of-the-envelope calculation indicates that the knowledge frontier effect accounts for 24% of the increase in team size. We calculate this as follows: team size in Soviet-rich fields increased by 33% from 1.34 to 1.78 in the before versus after period. We estimate that the Soviet-rich fields experience an 8% disproportionate increase (relative to Soviet-poor) during this period (Table 2, column 1). This represents 24% of the overall percentage increase. This is a rough calculation for many reasons not the least of which is that we assume that none of the increase in Soviet-poor subfields is due to an outward shift in the knowledge frontier, which is obviously an incorrect assumption since those fields experienced a normal flow of publications representing new knowledge throughout our forty year study period. However, since the calculation is also rough on many other dimensions we resist referring to our 24% estimate as a lower bound.

An important caveat is that we must exercise caution in generalizing the magnitude of our results in mathematics to other disciplines. Adams et al (2005) show that mathematics is somewhat of an outlier on at least two dimensions relative to other disciplines. First, it starts in 1981 with the smallest team size (1.53) relative to the major fields they study. The average team size across all fields is 2.77 and medicine has the largest (3.26). Second, it has the lowest annual growth rate during 1981-1990 (0.97). The average is 2.19 and physics has the highest (5.46). Similarly, mathematics has the second slowest growth rate in the 1990-1991 period (1.47) narrowly ahead of economics (1.29) and well below the average (2.57). Physics once again has the highest growth rate in this period (4.03) such that the average paper in physics has 7.26 authors by 1999. The increasing role of capital-intensive equipment (e.g., particle accelerators) is clearly a more salient factor in fields such as physics, engineering, and medicine than in mathematics. Therefore, even if 24% is a reasonable lower bound estimate of the fraction of the percentage increase in team size in mathematics that is likely an overestimate in fields where capital equipment plays a more central role.

We also find evidence that the knowledge frontier effect is related to the composition of research teams. The effect is realized strongly through teams that include a junior researcher on their team and is muted on teams that include a senior or star researcher. These results invite theory development to explain why the relative returns to collaborating seem to shift disproportionately for certain types of researchers. Perhaps the type of researchers that have the highest return from expanding their research team most value the type of skills those recently trained scholars provide. Perhaps juniors that are trained by recent Soviet émigrés have higher returns from collaborating due to their unusual distance from the frontier. Perhaps star researchers are able to assimilate new knowledge at the frontier in such a manner that the increase in their returns to collaborating are lower than average. Perhaps experienced researchers and stars make subtle shifts in their research trajectories to focus on topics that are least affected by the knowledge shock and thus have less benefit from collaborating. Our results on the frontier effect provide new insight to motivate theory development.

We must also exercise caution in extrapolating a general frontier effect from the effect of a shock. There exists a significant literature on the accumulation of knowledge. An underlying assumption of our interpretation of our estimates is that the team size response to a shock is similar to that for a gradual outward shift in the knowledge frontier. However, that may not be the case. Researchers may be able to absorb gradual increases in the knowledge frontier in a manner that does not generate as high returns to collaboration as those resulting from a sudden shock that may be more costly for researchers to internalize.

We began this paper by pointing out that unique policy implications associated with the frontier effect is one motivation for identifying the existence of the effect. In a series of papers (), Jones presents a variety of interventions that are potentially welfare enhancing in the presence of a frontier effect. These include subsidies and rewards to incentivize entry into careers in research, team-based evaluation of grant applications, and national or regional subsidies and specialization to prevent poverty traps due to underinvestment in human capital due to coordination failures arising from thin markets for complementary skills. While the evidence presented here has many limitations and caveats, it points to the possibility that the frontier effect is indeed non-trivial and worthy of further research.

References

(1) Adams, Black, Clemmons, Stephan 2005. Scientific Teams and Institutional Collaborations: Evidence from U.S. Universities, 1981-1999. Research Policy, Vol. 34 (3), pp. 259-285.

(2) Agrawal, A., Goldfarb, A. 2008. Restructuring Research: Communication Costs and the Democratization of University Innovation. American Economic Review, Vol. 98(4), pp. 1578–1590.

(3) Borjas, G. J., Doran, K. B. 2012. The Collapse of the Soviet Union and the Productivity of American Mathematicians. The Quarterly Journal of Economics, Vol. 127(3), pp. 1143-1203

(4) Dubois, P., Rochet, JC., Schlenker, M. 2011. Productivity and Mobility in Academic Research: Evidence from Mathematicians. Working paper.

(5) Ganguli, I. 2011. Saving Soviet Science: The Impact of Grants When Government R&D Funding Disappears. Working paper.

(6) Graham, L. R. 1993. Science in Russia and the Soviet Union: A short history. Cambridge University Press.

(7) Graham, L. R. , Dezhina, I. 2008. Science in the New Russia: Crisis, Aid, Reform. Indiana University Press.

(8) Grossman, J. W. 2002. The Evolution of the Mathematical Research Collaboration Graph. Congressus Numerantium, Vol. 158, pp. 201-212

(9) Hesse, Bradford W., Lee S. Sproull, Sara B. Kiesler, and John P. Walsh. 1993. "Returns to Science: Computer Networks in Oceanography," Communications of the ACM, 36(8): 90-101.

(10) Jones, B. 2009. The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder? Review of Economic Studies.

(11) Jones, B. 2010. As Science Evolves, How Can Science Policy? NBER Innovation Policy and the Economy, Vol. 11.

(12) Jones, B. 2011. The Knowledge Trap: Human Capital and Development, Reconsidered. NBER Working paper

(13) Kim, E.H., A. Morse, L. Zingales. 2009. Are elite universities losing their competitive edge? Journal of Financial Economics, Vol 93 (3), 353-381.

(14) Merton, R. K. 1973. The Sociology of Science: Theoretical and Empirical Investigations. University of Chicago Press, Chicago.

(15) Mowery, D., Nelson, R., Sampat, B., Ziedonis, A. 2004. Ivory Tower and Industrial innovation, Stanford University Press.

(16) Newman, M. E. J. 2004. Coauthorship Networks and Patterns of Scientific Collaboration. Center for the Study of Complex Systems and Department of Physics, University of Michigan, Ann Arbor.

(17) Oettl, Alex, Reconceptualizing Stars: Scientist Helpfulness and Peer Performance Management Science 58(6), pp. 1122–1140,

(18) Polyak, B.T. 2001. History of Mathematical Programming in the USSR: Analyzing the Phenomenon. Mathematical Programming.

(19) Rosen, S. 1981. The economics of superstars. Amer. Econom. Rev., 71(5) 845–858.

(20) Tikhomirov, V. M. 2007. On Moscow Mathematics - Then and Now. Golden Years of Moscow Mathematics, Second edition, co-publication of the AMS and the London Mathematical Society.

(21) Singh, J., L. Fleming. 2010. Lone inventors as sources of break-throughs: Myth or reality? Management Sci. 56(1) 41–56.

(22) Sonnenwald, D. H. 2007. Scientific Collaboration. Annual Review of Information Science and Technology, Vol 41, Issue 1, pp. 643-681.

(23) Stephan, P. 2012. How Economics Shapes Science. Harvard University Press.

(24) Waldinger, F. 2012. Peer effects in science—Evidence from the dismissal of scientists in Nazi Germany. Rev. Econom. Stud., Forthcoming.

(25) Walsh, J. P., Bayma, T. 1996. Computer Networks and Scientific Work. Social Studies of Science Vol. 26, Issue 3, pp. 661-703.

(26) Wuchy, S., Jones, B., Uzzi, B. 2007. The Increasing Dominance of Teams in Production of Knowledge. Science 316, p. 1036.

(27) Zucker, L. G., M. R. Darby, M. B. Brewer. 1998. Intellectual human capital and the birth of U.S. biotechnology enterprises. Amer. Econom. Rev. 88(1) 290–306.

Figure 1: Mathematics Taxonomy

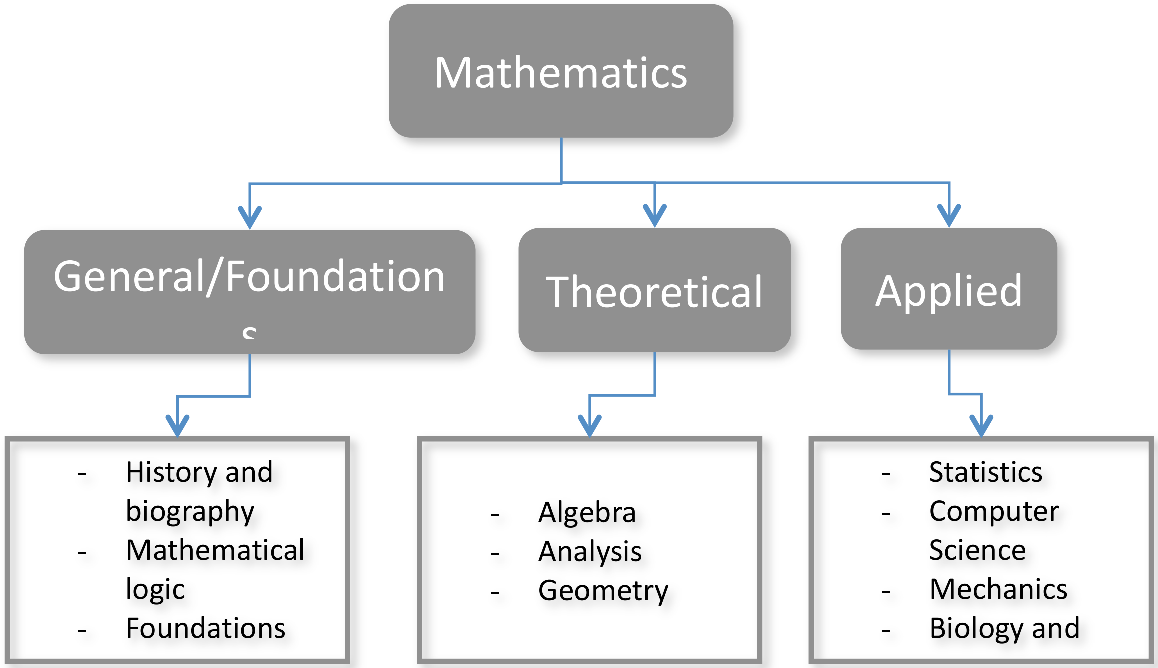
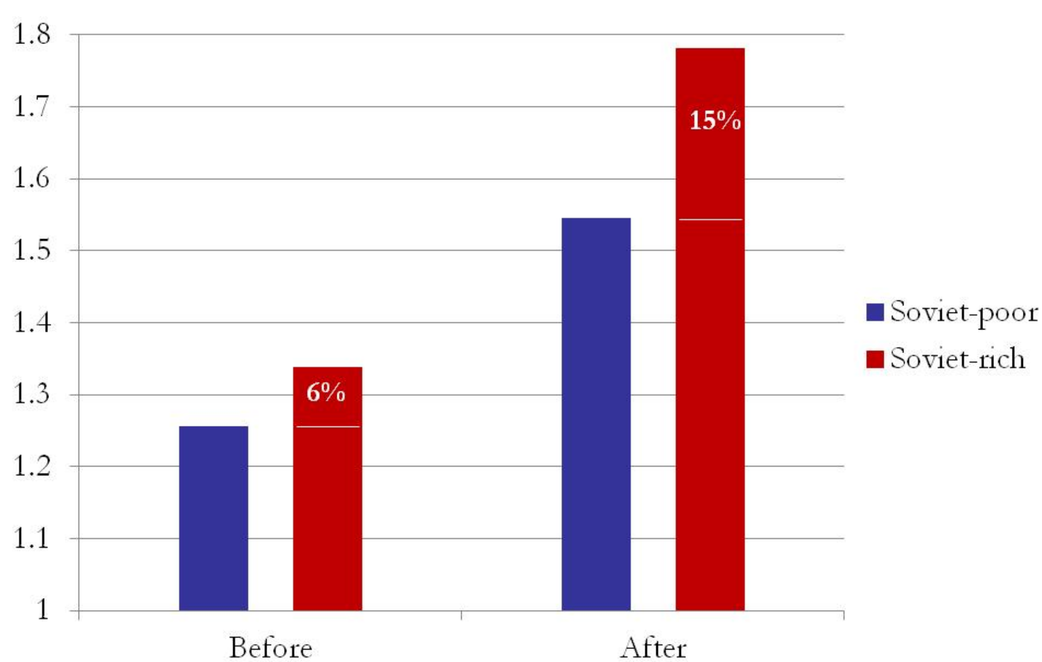
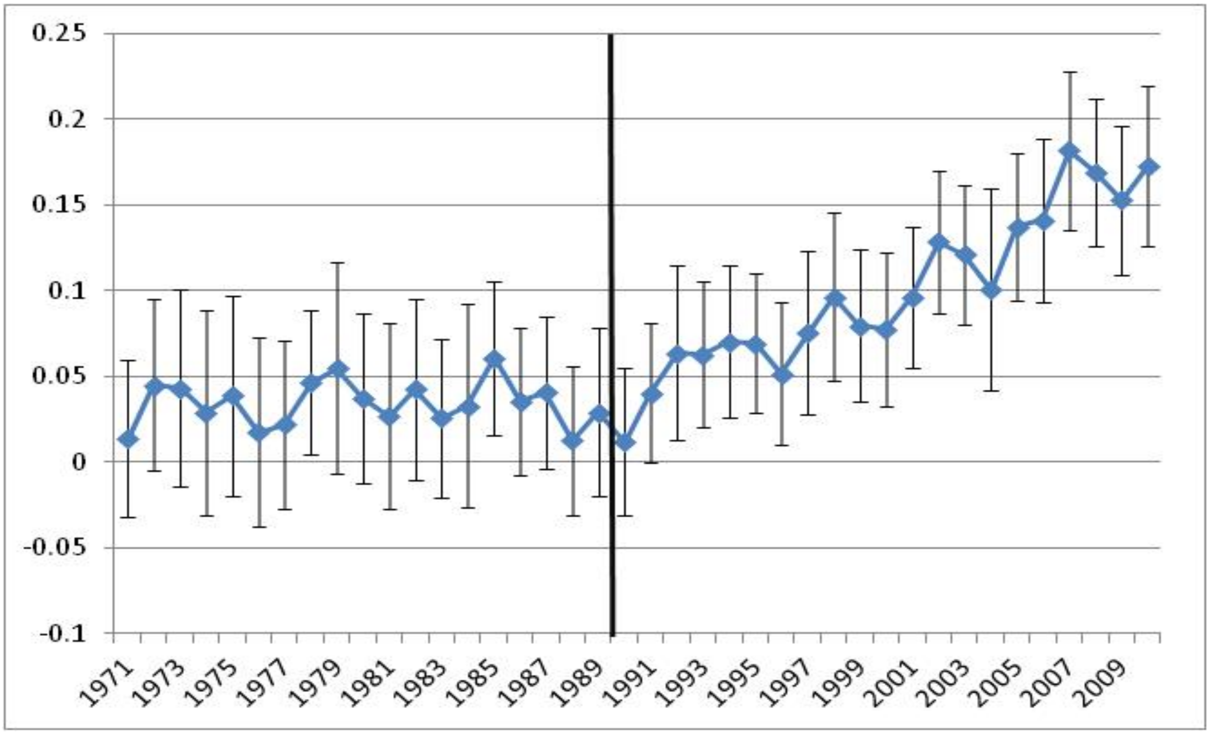


Figure 2: Disproportionate increase in mean team size in Soviet-rich subfields



Notes: This figure is based on 40 years of publication data for the three top and three bottom ranked subfields of theoretical mathematics. We plot the average team size for Soviet-rich and Soviet-poor subfields, before and after 1990.

Figure 3: Plot of estimated coefficients on interaction between Soviet-rich and year (DV = Team Size)



Notes: This figure is based on 40 years of publication data for the three top and three bottom ranked subfields of theoretical mathematics. Each point on the graph represents the coefficient value on the covariate Year*SovietRich and thus describes the relative difference in collaboration rates between Soviet-rich and -poor fields in that year. The bars surrounding each point represent the 95% confidence interval. All values are relative to the base year of 1970. We also present these results in table form in Appendix Table 2.

Table 1: Subfield rank based on proportion of publications (1984-1989) that are Soviet

Subfield Rank as per Borjas and Doran (2012)	MSC	Theoretical mathematics category	Description
1	45	Analysis	Integral equations
2	42	Analysis	Fourier analysis
3	35	Analysis	Partial differential equations
4	40	Analysis	Sequences, series, summability
5	31	Analysis	Potential theory
6	49	Analysis	Calculus of variations and optimal control; optimization
7	44	Analysis	Integral transforms, operational calculus
8	30	Analysis	Functions of a complex variable
9	8	Algebra	General algebraic systems
10	39	Analysis	Difference equations and functional equations
11	47	Analysis	Operator theory
12	17	Algebra	Non-associative rings and non-associative algebras
13	41	Analysis	Approximations and expansions
14	58	Geometry	Global analysis, analysis on manifolds
15	32	Analysis	Several complex variables and analytic spaces
16	31	Analysis	Special functions
17	22	Algebra	Topological groups, lie groups, and analysis upon them
18	54	Geometry	General topology
19	20	Algebra	Group theory and generalizations
20	28	Algebra	Measure and integration
21	18	Algebra	Category theory; homological algebra
22	55	Analysis	Algebraic topology
23	26	Algebra	Real functions, including derivatives and integrals
24	52	Geometry	Convex geometry and discrete geometry
25	14	Algebra	Algebraic geometry
26	43	Analysis	Abstract harmonic analysis
27	15	Algebra	Linear and multilinear algebra; matrix theory
28	6	Algebra	Order theory
29	12	Algebra	Field theory and polynomials
30	5	Algebra	Combinatorics
31	51	Geometry	Geometry
32	57	Geometry	Manifolds
33	13	Algebra	Commutative rings and algebras

Notes: This ranking is adapted from Borjas and Doran (2012). It is based on the ratio of the number of Soviet versus American papers published in the particular subfield between 1984 and 1989. Papers are defined as Soviet if at least one author has a Soviet institutional affiliation. American papers are defined similarly.

Table 2: Teams in Soviet-rich subfields exhibit disproportionate increase in team size after 1990

Dependent variable: log of author count				
	Soviet-rich: top three	Soviet-rich: top five	Soviet-rich: top 25%	Soviet-rich: top 50%
AfterIronCurtain SovietRich*	**0.0785***** (0.0033)	**0.0252**** (0.0070)	**0.0360***** (0.0048)	**0.0156**** (0.0048)
Year fixed effects	Yes	Yes	Yes	Yes
Subfield fixed effects	Yes	Yes	Yes	Yes
R-squared	0.113	0.122	0.107	0.106
Observations	133,497	204,470	563,462	563,462
	The unit of analysis is the publication. All models are OLS with robust standard errors, clustered around treated/not and before/after *significant at 10%, **significant at 5%, ***significant at 1%			

Table 3: Teams in Soviet-rich subfields exhibit disproportionate increase in propensity to cite Soviet prior art after 1990

Dependent variable: References to Soviet art				
	Count of Soviet references (by name)	Percentage of Soviet references (by name)	Count of Soviet references (Soviet journal)	Percentage of Soviet references (Soviet journal)
*SovietRich * AfterIronCurtain*	**0.4481**** (0.0295)	**0.0133***** (0.0019)	**0.4265***** (0.0299)	**0.0193**** (0.0009)
Year fixed effects	Yes	Yes	Yes	Yes
Subfield fixed effects	Yes	Yes	Yes	Yes
R-squared	0.170	0.125	0.136	0.101
Observations	743	743	743	743
	The unit of analysis is the publication. The sample includes the top/bottom three subfields drawn from the top 30 journals three years before and after the collapse of the Soviet Union. All models are OLS with robust standard errors, cluster around treated/not and before/after *significant at 10%, **significant at 5%, ***significant at 1%			

Table 4: At least some of the knowledge frontier effect occurs on the extensive margin

Dependent variable: dummy variable equal 1 for multi-author papers and 0 otherwise				
	Soviet-rich: top three	**Soviet-rich: top five**	**Soviet-rich: top 25%**	**Soviet-rich: top 50%**
AfterIronCurtain* SovietRich	**0.0822*** (0.0030)**	**0.0430*** (0.0056)**	**0.0403*** (0.0042)**	**0.0232** (0.0042)**
Year fixed effects	Yes	Yes	Yes	Yes
Subfield fixed effects	Yes	Yes	Yes	Yes
R-squared	0.101	0.106	0.093	0.093
Observations	133,497	204,470	563,462	563,462
	All models OLS with robust standard errors, clustered around Soviet rich/poor and before/after *significant at 10%, **significant at 5%, ***significant at 1%			

Table 5: Intensive Margin

Dependent variable: log of author count; sample conditioned on multi-author papers				
	Soviet-rich: top three	**Soviet-rich: top five**	**Soviet-rich: top 25%**	**Soviet-rich: top 50%**
AfterIronCurtain* SovietRich	**0.0098** (0.0012)**	**-0.0089** (0.0018)**	**0.0074** (0.0020)**	**-0.0090** (0.0023)**
Year fixed effects	Yes	Yes	Yes	Yes
Subfield fixed effects	Yes	Yes	Yes	Yes
R-squared	0.034	0.047	0.041	0.041
Observations	59,375	94,614	236,965	236,965
	All models OLS with robust standard errors, clustered around Soviet rich/poor and before/after *significant at 10%, **significant at 5%, ***significant at 1%			

Table 6: Junior scholars play an important role in the knowledge frontier effect

Dependent variable: Log of author count per publication per year				
	Age = 1 if at least one author in the youngest 10%	**Age = 1 if at least one author in the youngest 25%**	**Age = 1 if at least one author in the oldest 10%**	**Age = 1 if at least one author in the oldest 25%**
SovietRich * AfterIronCurtain* Age	**0.0476*** (0.0014))**	**0.0605*** (0.0012)**	**-0.0157** (0.0032)**	**-0.0018 (0.0041)**
Age	0.0861*** (0.0036)	0.0817*** (0.0036)	0.1683*** (0.0030)	0.1661*** (0.0020)
*SovietRich*Age*	0.0252*** (0.0016)	0.0267*** (0.0013)	0.0083*** (0.0011)	0.0121*** (0.0006)
*AfterIronCurtain*Age*	0.0411*** (0.0035)	0.0317*** (0.0037)	0.0254** (0.0058)	0.0860*** (0.0046)
AfterIronCurtain SovietRich*	0.0698*** (0.0030)	0.0644*** (0.0028)	0.0876*** (0.0035)	0.0740*** (0.0027)
Year fixed effects	Yes	Yes	Yes	Yes
Subfield fixed effects	Yes	Yes	Yes	Yes
R-squared	0.135	0.137	0.135	0.177
Observations	133,835	133,835	133,835	133,835
	All models are OLS with robust standard errors, clustered around Soviet rich/poor and before/after Top and bottom 3 ranking subfields of theoretical mathematics *significant at 10%, **significant at 5%, ***significant at 1%			

Table 7: Teams with a research star play a lesser role in the knowledge frontier effect

Dependent variable: Log of author count per publication per year		
	At least one author is in Top 10%	**At least one author is in Top 25%**
SovietRich * AfterIronCurtain* Star	**-0.0450*** (0.0015)**	**-0.0309*** (0.0028)**
Star	0.1402*** (0.0011)	0.1262*** (0.0007)
*SovietRich*Star*	0.0309 (0.0010)	0.0321 (0.0004)
*AfterIronCurtain*Star*	0.0871*** (0.0006)	0.1070*** (0.0010)
AfterIronCurtain SovietRich*	0.0855*** (0.0021)	0.0820*** (0.0010)
Year fixed effects	Yes	Yes
Subfield fixed effects	Yes	Yes
R-squared	0.148	0.165
Observations	133,835	133,835
	All models are OLS with robust standard errors, clustered around Soviet rich/poor and before/after Top and bottom 3 ranking subfields of theoretical mathematics *significant at 10%, **significant at 5%, ***significant at 1%	

Appendix Table 1 – Robustness check for Table 2

Dependent variable: log of author count per publication per year				
	Soviet-rich: top three	Soviet-rich: top five	Soviet-rich: top 25%	Soviet-rich: top 50%
AfterIronCurtain SovietRich*	**0.0805***** (0.0032)	**0.0672**** (0.0040)	**0.0359***** (0.0042)	**0.0130*** (0.0048)
Year fixed effects	Yes	Yes	Yes	Yes
Subfield fixed effects	Yes	Yes	Yes	Yes
R-squared	0.109	0.109	0.107	0.106
Observations	131,061	190,473	563,462	563,462
	All models are OLS with robust standard errors, clustered around Soviet rich/poor and before/after Here we use our own ranking scheme rather than Borjas and Doran (2012); our scheme uses worldwide publications from 1970-1989 rather than US publications from 1984-1989 but relies on identifying Soviet publications by name rather than affiliation data *significant at 10%, **significant at 5%, ***significant at 1%			

Appendix Table 2 – Coefficient estimates used to plot Figure 3

Dependent variable: Log of author count per publication per year	
	Soviet-rich: top three
*SovietRich*1971*	0.0137 (0.0233)
*SovietRich*1972*	0.0452* (0.0256)
*SovietRich*1973*	0.0431 (0.0292)
*SovietRich*1974*	0.0290 (0.0305)
*SovietRich*1975*	0.0389 (0.0296)
*SovietRich*1976*	0.0174 (0.0280)
*SovietRich*1977*	0.0219 (0.0248)
*SovietRich*1978*	0.0465** (0.0212)
*SovietRich*1979*	0.0550* (0.0315)
*SovietRich*1980*	0.0375 (0.0253)
*SovietRich*1981*	0.0269 (0.0273)
*SovietRich*1982*	0.0425 (0.0269)
*SovietRich*1983*	0.0258 (0.0235)
*SovietRich*1984*	0.0330 (0.0303)
*SovietRich*1985*	0.0607*** (0.0226)
*SovietRich*1986*	0.0356 (0.0217)
*SovietRich*1987*	0.0408* (0.0225)
*SovietRich*1988*	0.0127 (0.0222)
*SovietRich*1989*	0.0291 (0.0249)
*SovietRich*1990*	0.0119 (0.0218)
*SovietRich*1991*	0.0405* (0.0208)
*SovietRich*1992*	0.0639** (0.0257)

*SovietRich*1993*	0.0628*** (0.0215)
*SovietRich*1994*	0.0702*** (0.0225)
*SovietRich*1995*	0.0694*** (0.0207)
*SovietRich*1996*	0.05160** (0.0211)
*SovietRich*1997*	0.0754*** (0.0241)
*SovietRich*1998*	0.0965*** (0.0248)
*SovietRich*1999*	0.0797*** (0.0225)
*SovietRich*2000*	0.0778*** (0.0229)
*SovietRich*2001*	0.0963*** (0.0208)
*SovietRich*2002*	0.1287*** (0.0212)
*SovietRich*2003*	0.1213*** (0.0206)
*SovietRich*2004*	0.1007*** (0.0301)
*SovietRich*2005*	0.1376*** (0.0218)
*SovietRich*2006*	0.1415*** (0.0242)
*SovietRich*2007*	0.1818*** (0.0233)
*SovietRich*2008*	0.1691*** (0.0219)
*SovietRich*2009*	0.1528*** (0.0222)
*SovietRich*2010*	0.1729*** (0.0237)
Year fixed effects	Yes
Subfield fixed effects	Yes
R-squared	0.114
Observations	133,497
	All models are OLS with robust standard errors, clustered around subfield and year *significant at 10%, **significant at 5%, ***significant at 1%