The Collapse of the Soviet Union and the Productivity of American Mathematicians
Faculty Research Working Paper Series

George J. Borjas
Harvard Kennedy School and NBER

Kirk B. Doran
University of Notre Dame

THE COLLAPSE OF THE SOVIET UNION AND THE PRODUCTIVITY OF AMERICAN MATHEMATICIANS

George J. Borjas
Kirk B. Doran

The Collapse of the Soviet Union and the Productivity of American Mathematicians
George J. Borjas and Kirk B. Doran
NBER Working Paper No. 17800
February 2012
JEL No. J61,O31

<u>ABSTRACT</u>

It has been difficult to open up the black box of knowledge production. We use unique international data on the publications, citations, and affiliations of mathematicians to examine the impact of a large post-1992 influx of Soviet mathematicians on the productivity of their American counterparts. We find a negative productivity effect on those mathematicians whose research overlapped with that of the Soviets. We also document an increased mobility rate (to lower-quality institutions and out of active publishing) and a reduced likelihood of producing "home run" papers. Although the total product of the pre-existing American mathematicians shrank, the Soviet contribution to American mathematics filled in the gap. However, there is no evidence that the Soviets greatly increased the size of the "mathematics pie." Finally, we find that there are significant international differences in the productivity effects of the collapse of the Soviet Union, and that these international differences can be explained by both differences in the size of the émigré flow into the various countries and in how connected each country is to the global market for mathematical publications.

George J. Borjas
Harvard Kennedy School
79 JFK Street
Cambridge, MA 02138
and NBER
gborjas@harvard.edu

Kirk B. Doran
438 Flanner Hall
University of Notre Dame
Notre Dame, IN 46556
kdoran@nd.edu

**The Collapse of the Soviet Union
and the Productivity of American Mathematicians**[*]

George J. Borjas and Kirk B. Doran

I. Introduction

Many economists believe that knowledge production is central to long-term economic growth. Nevertheless, it has been difficult to document the factors that enter the production function of knowledge. The difficulty arises for many reasons. For example, knowledge production is both social and reciprocal (Lucas 2009), in that the ideas of one researcher influence and are, in turn, influenced by the ideas of others. Similarly, basic knowledge production can be difficult to observe and measure at the individual level, so that we do not have a well-developed set of facts that can help guide our thinking on a theoretical framework. Finally, we do not understand why knowledge in some fields progresses at a remarkable rate in a short-lived burst, yet stagnates for decades in other areas.[1]

Despite these difficulties, it is clear that there are countless possibilities for spillovers when producing knowledge: the knowledge produced by one researcher is both

[1] Jones (2005, p. 1107) succinctly summarizes the difficulties: "While we have made much progress in understanding economic growth in a world where ideas are important, there remain many open, interesting research questions. The first is, 'What is the shape of the idea production function?' How do ideas get produced?...The current research practice of modeling the idea production function as a stable Cobb-Douglas combination of research and the existing stock of ideas is elegant, but at this point we have little reason to believe that it is correct." See also Lucas (1988), Romer (1986), Romer (1990), and Jones and Romer (2010).

an output and an input into another researcher's production function. As Newton said, "If I have seen further it is by standing on the shoulders of giants."

But even if the ideas of a highly qualified single worker spill over to other workers with whom they interact, the overall effect of the interaction can still be deleterious to the productivity of other workers. In particular, in a world with constraints on the funding and dissemination of ideas (e.g., a limit on the number of faculty slots, or, more abstractly, a limit on the attention span of the potential audience), large and sudden increases in the population of producers of knowledge can result in diminishing marginal productivity for a pre-existing worker. As an example, a young academic might appreciate the hiring of a new illustrious colleague in, say, mathematics because it may improve his own ideas. At the same time, the young academic realizes that in a world with limited funding and limited research opportunities, his own services and research now become relatively less important to his own department and to the field in general.

A number of recent empirical studies have attempted to quantify the net impact of the presence of a highly skilled worker on the ideas and output of other workers. For example, Waldinger (2009) examines the productivity of the doctoral students who were left behind when superstar German scientists left Germany during the Nazi era. He finds that these students suffered in the absence of their highly skilled mentors. In more recent work, however, Waldinger (2011) finds that the coauthors left behind in Nazi Germany did not experience a loss in productivity when the superstars left. Finally, Azoulay, Zivin, and Wang (2010) document the decreased output suffered by the co-authors of superstar scientists after the superstars die. They find that the coauthors become much less productive when the superstar is no longer able to collaborate. In concluding, they note:

"Although we measure the impact of losing a star collaborator, a full accounting of knowledge spillovers would require information on the benefits that accrued to the field while the star was alive. We can think of no experiment, natural or otherwise, that would encapsulate this counterfactual" (Azoulay, Zivin, and Wang 2010, p. 580).[2]

This paper attempts to measure the productivity effects of the entry of highly skilled scientists in a context where we can observe the counterfactual of no entry. In particular, we examine the impact of the influx of renowned Soviet mathematicians into the global mathematics community. In the period between the establishment and fall of communism, Soviet mathematics developed in an insular fashion and along very different specializations than American mathematics. As a result, some mathematicians experienced few potential insights from Soviet mathematics after the collapse of the Soviet Union, while other fields experienced a flood of new mathematicians, theorems, and ideas.

We have constructed a data set that contains information on the authorship of every paper published in mathematics over the past 70 years.[3] These data allow us to document the location, affiliation, and complete publication and citation records of mathematicians who were active in the Soviet Union and around the world for the past few decades.

Prior to the collapse of the Soviet Union, there was little collaboration and only infrequent exchanges between Soviet and Western mathematicians. In fact, every written

---

[2] Related research on knowledge production includes Furman, Kyle, Cockburn, and Henderson (2005) and Hunt and Gauthier-Loiselle (2010).

[3] Mathematical research plays a fundamental role in technological progress. In fact, recent applications of modern academic research papers in mathematics and related fields (such as theoretical computer science and mathematical physics) to our broader economy are so numerous and diverse that it is impossible to characterize them briefly. For a few examples, consider the Rivest-Shamir-Adleman algorithm that forms the backbone of Internet encryption; the Reed–Solomon error correction that makes possible compact disks, deep-space communication, error-free bar codes, and DSL television; and Claude Shannon's information theory, which has been applied everywhere from gambling and investing, to bioinformatics and music, and even to the discovery of new oil fields with seismic oil exploration.

communication with an American mathematician was opened and read by the authorities, and special permission was required to publish outside the Soviet Union. Depending on the era in the Cold War, Soviet violators could be imprisoned (Polyak 2002). After the collapse of the Soviet Union, over 1,000 Soviet mathematicians migrated to other countries, with a large fraction settling in the United States. In addition, the mathematicians who remained in the Soviet Union became part of the globalized publications market in mathematics.[4]

Our empirical analysis demonstrates that the American mathematicians whose research programs most overlapped with that of the Soviets experienced a reduction in productivity after the entry of Soviet émigrés into the U.S. mathematics market. This effect is observed at both tails of the quality distribution of mathematicians. First, the likelihood of a competing mathematician producing a "home run" paper fell significantly. Similarly, marginal mathematicians became much more likely to move to lower-quality institutions and to exit knowledge production altogether. We also find evidence that the students of the Soviet émigrés had higher lifetime productivity than other students from the same institution who had non-émigré advisors. However, this gain was more than offset by the productivity loss suffered by students who had American advisors with Soviet-like research programs.

On aggregate, based on the pre-1992 age-output profile of American mathematicians, we find that the actual output of mathematicians with a Soviet-like research program is far below what one would have expected. Even though there was a net

---

[4] Abramitzky and Sen (2011) provide a detailed analysis of how the collapse of Soviet communism led to a diffusion of knowledge to and from the West and the former Soviet Union.

loss in total output for American mathematicians, this loss was approximately made up by the published output of the Soviet émigrés in the United States.

Our analysis also directly confronts a related issue in the knowledge production literature. Knowledge producers interact in both a job market (i.e., the market for selling human capital), and in a publications market (i.e., the market for selling codified knowledge, such as journal articles). The collapse of the Soviet Union provides a unique opportunity to empirically examine the relative importance of each market for determining the net productivity effect on knowledge producers.

After all, the collapse of the Soviet Union induced many Soviet mathematicians to move to some countries, but not to others, creating competitive pressures in some job markets, but not in others. At the same time, the market for journal space has some segments that are country-specific, but other segments that cross over geographic boundaries. As a result, mathematicians in countries that did not physically receive many Soviet émigrés may have experienced competition in the journal market, while mathematicians in still other countries did not experience increased competition in either market. The international differences in the post-1992 productivity of mathematicians reveals that competitive pressures in both the job market and in the market for codified knowledge are important determinants of productivity and of crowd-out effects.

II. Historical Context

After the establishment of the Soviet Union in 1922, Soviet mathematics entered a long period of development independent from mathematics in the West. To varying degrees between 1922 and 1992, the Soviet Government instituted strict controls on which

scientists could communicate with Western peers, on the parameters of scientific travel, on the acceptable outlets for publication, and on access to Western materials.[5] Just as speakers of one language, when separated geographically for many generations, eventually develop separate and different dialects through natural changes over time, so Western and Eastern mathematicians, separated by Stalinist and Cold War political institutions, developed under different influences to the point of achieving very different specializations across the fields of mathematics.

An important event that cemented the isolation of Soviet mathematicians was the "Luzin affair." In 1936, Nikolai Luzin, a mathematician at Moscow State University and a member of the USSR Academy of Sciences, became the target of a political campaign. The allegations included not only the usual charge of promoting anti-Soviet propaganda, but also the specific accusation that Luzin saved his main academic results for publications in foreign outlets. Although Stalin eventually spared Luzin's life, the impact on Soviet mathematics was swift and dramatic: "The main visible consequence of the Luzin affair was

---

[5] Polyak (2002, p. 2) gives a firsthand account of the life of mathematicians behind the Iron Curtain. Polyak writes: "'The iron curtain' was not only a metaphor, it was a real obstacle to international contacts...When Professor Ya.Z. Tsypkin received a letter in the late 1940s from an American reader of his paper, he was summoned by the KGB and underwent a long investigation there, tottering at the edge of arrest. . . Another source of difficulties for researchers was the mania for secrecy. . . Nobody was allowed to publish any paper without special permission confirming that the publication does not contradict numerous security restrictions. All letters abroad (as well as letters from abroad) were opened and inspected. Everybody must have special permission, and a full text of the talk had to be approved if you were going to an international conference. And working in a classified institution (which was the case for many experts in mathematical programming), complicated the situation drastically. . . The situation in the 1940s to mid-1950s was the worst. Malevolent intent by the authorities could lead a researcher to the GULAG. The period 1955–1970 was the least oppressive, it was a 'golden age' of Soviet mathematics. . . The years 1970 to 1985 were a period of stagnation in political, social, economic, and scientific life. All the troubles I have mentioned above played a more and more significant role in the development of Soviet science and thus led to its degradation."

that, from this precise moment, Soviet mathematicians began to publish almost exclusively

in Soviet journals and in Russian."[6]

Figure I illustrates the dramatic differences in specializations between Soviet and

American mathematics. In particular, the figure gives the number of papers published

before 1990 by Soviet mathematicians in a particular field relative to the number of papers

published by American mathematicians in that field. Despite the fact that, on aggregate,

Americans published about three times as many papers as the Soviets, Soviet

mathematicians published 1.4 papers per American paper in Integral Equations. In

contrast, Soviet mathematicians published only 0.06 papers per American paper in

Statistics. In fact, the two most popular Soviet fields were Partial Differential Equations and

Ordinary Differential Equations, and these two fields accounted for 17.8 percent of all

publications. In contrast, the two most popular American fields were Statistics and

Operations Research, Mathematical Programming, and these two fields accounted for 15.6

percent of all American publications.[7]

The primary influences behind the development of both Soviet and Western

mathematics were history dependence and, to a lesser extent, state funding. In the Soviet

Union, for example, the mathematical genius Andrey Kolmogorov developed important

results in the area of Probability and Stochastic Processes beginning in the 1930s. In a

scenario common throughout Soviet mathematical history, he then established a "school"

---

[6] O'Connor and Robertson (1999). The charges against Luzin were published in *Pravda* in a series of articles beginning on July 3, 1936. The Luzin affair involved intrigue, blackmail, and betrayal—even Andrey Kolmogorov played a non-trivial role in the play (Graham and Kantor 2009); see also Smithies (2003).

[7] The third most popular field, Quantum Theory, was the same in both countries, accounting for 6.8 percent and 5.9 percent of publications in the USSR and the United States, respectively. The fourth and fifth most popular fields in the Soviet Union were Probability Theory and Stochastic Processes (6.1 percent) and Global Analysis, Analysis on Manifolds (4.8 percent). The respective fields in the United States were Computer Science (4.5 percent) and Numerical Analysis (4.3 percent).

at Moscow State University, attracting some of the best young minds over the next four decades, such as the teenage prodigy Vladimir Arnold in the 1950s. Arnold himself quickly solved Hilbert's famous "Thirteenth Problem," and initiated the field of symplectic topology. The large amount of Soviet work in these areas even in the 1980s can be seen in Figure I under the subject headings "Mechanics of Particles and Systems," "Global Analysis, Analysis on Manifolds," and "Ordinary Differential Equations." Since the United States did not have the unique Kolmogorov-Arnold combination, the amount of work done by American mathematicians in these subfields was far less than would have been expected given the size and breadth of the American mathematics community.

In the United States, however, researchers like John Milnor at Princeton University and Raoul Bott at Harvard University developed key ideas in the topology of manifolds in the 1950s. Their students and collaborators produced an enviable body of research in Manifolds and Cell Complexes, which, because of the lack of a similar chance originator in their country, was never replicated in the Soviet Union.[8]

Finally, Soviet funding was limited (in comparison to the U.S.) in fields requiring experiments or equipment (Howe 1990). Figure I also demonstrates that as late as the 1980s, this resulted in a large discrepancy between American and Soviet specialization in computer science and related fields.[9]

---

[8] Algebraic Geometry, a field of relative U.S. excellence, provides another example of the persistence of history dependence. The Summary Report of the Panel on Soviet Mathematics (Lefschetz 1961, p. IV-2) explains that: "in no part of mathematics is the Soviet Union weaker than in algebraic geometry. No significant contributions have ever . . . come from there." This was not only true in 1961, but, as Figure I demonstrates, it remained true in the 1980s.

[9] Personal communication with Professor Lawrence Shepp, who during a long career at Bell Labs was in contact with some Soviet mathematicians as early as 1964, suggests that the fact that Soviet mathematicians avoided the most popular American field (Statistics) had little to do with funding. Specializing in statistics was politically dangerous in the USSR, as it would require a great deal of massaging of sensitive data.

We exploit the fact that Soviet and American mathematicians did not choose their specializations in the 1970s and 1980s in the belief that they would soon have an opportunity to co-author papers, compete for pages in the same journals, and apply for jobs at the same universities. The consensus among both Soviet and American experts almost immediately before the collapse of communism was that the political system of the existing Soviet state was not ripe for a sudden change. Walter Laqueur (1996, p. 65) describes how in the Soviet Union itself, "most believed the system was so strong that it would never essentially change. Others, more optimistic, thought that change was perhaps possible over a long period—decades, or more likely, generations." In the West, Laqueur (1996, p. 99) reports that Sovietologists were taken by surprise: "The U.S. government (like most others) had enormously overrated Soviet economic performance. . .According to a study published as late as 1988 by a well-known Western economist specializing in the Soviet Union, Soviet citizens enjoyed 'massive economic security' . . . the consensus was that the Soviet Union was *not* on the verge of economic bankruptcy and political disintegration."

Thus, the divergent interests and capabilities of Soviet and American mathematics that had emerged in earlier decades were not likely to have been modified in the 1980s by any serious belief that the Soviet isolation would soon end. As Figure II shows, the negligible pre-1990 rate of co-authorship between mathematicians reporting Soviet research addresses and mathematicians reporting U.S. addresses does not engender the hope that the number of such collaborations would suddenly explode.

Around 1990, as the political situation changed in the Soviet Union, a large number of Soviet mathematicians began to come into regular contact with Western mathematicians through visits and immigration. According to American mathematicians who witnessed this

sudden increase in contact opportunities, the effect on American mathematics was immediate. In 1990, the *New York Times* reported (Kolata 1990):

> "American scientists say they have benefited immensely from the [recent] Soviet visitors. . . 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 [Soviet mathematician] 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 . . . No one could help. But . . . Soviet scientists had done a lot of work on such problems. 'It was a whole new world I had access to,' Dr. Diaconis said. 'Together, we'll be able to solve the problem.'"

Inevitably, the American mathematical community also experienced increased competition in hiring.[10] The American Mathematical Society's 1991-1992 *Academic Hiring Survey* reports that: "Citizens of Eastern European countries and the former Soviet Union accounted for 13% of all newly-hired faculty and 15% of the tenured and tenure-eligible new hires" (McClure 1992, p. 311). The report also identifies "increased numbers of highly qualified recent U.S. immigrants seeking employment in academia" as a leading cause of the unprecedented 12 percent unemployment rate of new American mathematics Ph.Ds. in 1991 (McClure 1992, p. 312).[11] Figure III illustrates the employment trends of newly minted doctorates from North American institutions. It is evident that there was a dramatic increase in the unemployment rate (as well as a dramatic decrease in the probability of

---

[10] For example, Soviet mathematician Nicolai Reshetikhin not only helped Professor Diaconis with his puzzle, but also accepted a job as a Visiting and Assistant Professor at Harvard from 1989 through 1991.

[11] At the same time that the unemployment rate of newly minted mathematicians was rising rapidly in the early 1990s, the economy-wide unemployment rate for college graduates was falling from 3.2 to 2.2 percent between 1992 and 1996.

obtaining a position in research universities) at the same time that the Soviet influx was

occurring.[12]

It is clear that the ability to communicate and collaborate on a one-to-one basis (and

especially emigrate) increased greatly after the collapse of the Soviet Union. As Diaconis

discovered, ideas and knowledge became much easier to share. The economics of

innovation literature distinguishes between two broad types of knowledge: codified

knowledge and tacit knowledge. Codified mathematical knowledge is recorded in journal

articles and books, while tacit knowledge is part of the un-codified human capital of

mathematical practitioners.[13]

Both codified and tacit mathematical knowledge were costly to share during the

Soviet era, even though translations of Soviet papers in academic journals were often

available.[14] Moreover, some of the codified knowledge remained un-translated. Soviet

scientists often published their best work in books: "The Russians are prolific producers of

books . . . and the consensus of experts is that the best Russian authors reserve some of

their best work for books which often constitute first publication of important research

---

[12] The exodus of key scientific personnel from the former Soviet Union to the West led George Soros to establish a program that provided research funds to those scientists who chose to remain; see Ganguli (2010) for an analysis of the impact of this program on career choices.

[13] "Typically, new knowledge and expertise have a broad tacit dimension, meaning that they are neither articulated nor codified. Tacit knowledge resides in people, institutions, or routines" (Foray 2004, pp. 17-18).

[14] The National Science Foundation, American scientific societies, and private businesses had begun translating Soviet scientific journals from Russian into English as far back as 1949 (O'Dette 1957). It would seem, therefore, that the collapse of the Soviet Union represents but a trivial change in the cost of obtaining access to the specific type of codified knowledge contained in journal articles. But the citations results described below suggest that Americans may not have read as much Soviet work as would be expected, and history records famous examples of American unfamiliarity with important Soviet journal articles that were available during the Cold War: "...Even when high-quality translations are readily available, the information can still be lost if they are not read. . . . In January 1962, the National Aeronautics and Space Administration announced that Explorer XII, a satellite launched in August 1961, had discovered a new radiation belt with an 'unexpected boundary' at 40,000 miles from the earth. In fact, the same belt was discovered by Soviet rockets in 1959 and this was reported in translations of *Soviet Physics Dolkady* in 1960 and *Soviet Astronomy AJ* in 1961" (Tybulewicz 1970, pp. 55-56).

information" (O'Dette 1957, p. 581). But even though mathematics books contained some

of the most important Soviet results, Abramitzky and Sen (2011) report that the translation

rate of hard-science Eastern Bloc books into English was extremely low. After the collapse

of the Soviet Union, tacit mathematical knowledge and an easy introduction to the codified

knowledge were as close as a phone call, an unobstructed correspondence, or even a knock

on the office next door.[15]

Evidence of this knowledge shock is apparent in the citation trends of American

mathematicians to Soviet and American articles. Figures IVa and IVb show the number of

citations that American mathematicians made to other American papers during the period,

as well as the citations Americans made to Soviet papers (defined as papers published by

an author using a Soviet affiliation).[16] Both before and after the collapse of the Soviet

Union, most American mathematicians exhibit "home bias," preferring to cite results in

other American papers. Considering that the total number of American papers was only

three times that of the Soviet Union, this home bias is extreme. Nevertheless, there is a

substantial increase in the number of citations to Soviet papers after 1990. In fact, the

share of American citations to Soviet research approximately tripled. It is remarkable that

---

[15] The inadequacy of the international phone infrastructure during the Soviet era is legendary, making scientific collaboration harder, but undoubtedly making the KGB's job of hampering such collaboration easier: "The international telephone transit center in Moscow . . . has a capacity of 1,500 circuits . . .This exchange has to handle all the international calls from the RSFSR [Russian Republic], as well as calls in transit from the other republics and from a number of Eastern European and other socialist countries. Since it allows a maximum of around 800 outgoing calls at any one time, it constitutes a major bottleneck in communications between the USSR and the rest of the world. Moreover, most circuits are dedicated to links with Eastern European and other socialist countries; there are, for example only 40 circuits for calls to the United Kingdom, 25 to France and 16 to the United States" (IMF 1991, p. 128).

[16] Since an author's geographic location is only rarely available before 1978, we construct the statistic as follows. The numerator is the number of citations to American (or Soviet) papers written between 1978 and year t made by American papers written between 1978 and t. The denominator is the total number of citations to papers with a valid research address written between 1978 and t made by American papers written during that time.

Soviet papers should become much more cited since many of the Soviet Union's best

mathematicians had emigrated, and many of them were in the United States. It is equally

remarkable that there was also an increase in citations to "old" Soviet work (i.e., papers

written before the fall of communism). Relative to its pre-treatment trend, the citation

share of "old" Soviet papers out of the citations to all papers became 40 percent higher

after the fall.[17]

In sum, the collapse of the Soviet Union was associated with a decrease in the cost of

accessing both codified and tacit knowledge, and the resulting diaspora of Soviet

mathematicians after 1992 led to both a mathematical labor supply shock and a

mathematical knowledge shock in many countries.

III. Data and Summary Statistics

Our data are drawn from three distinct, but related sources. The Supplementary

Data Appendix gives a more detailed description of the data sets and the construction of

the samples.

First, the American Mathematical Society (AMS) provided us with a database that

reports the number of papers published by every mathematician, by field and year.[18] In

addition to the information on the number of papers, the AMS data reports the

---

[17] To calculate the frequency of citations made by American papers to pre-1990 Soviet papers, we calculated the following fraction: the denominator is citations made by American papers published each year to any papers with identifiable research addresses written between 1978 and 1989; the numerator is the number of such citations to papers that had at least one author in the Soviet Union. We then estimated a simple regression of the log citation share on: year, year squared, and a post-1990 indicator variable. The model's fit is improved substantially by the inclusion of the post-1990 dummy variable, and its coefficient implies a 40 percent increase in American attention to pre-1990 Soviet work after the collapse of the Soviet Union,

[18] The AMS file provided to us contains information for all mathematicians who have published at least one paper since 1939.

mathematician's institutional affiliation at the time the paper was published, as well as the location of the affiliation. The AMS, however, only began to collect the affiliation information on a systematic basis around 1984, so that affiliation and location are not typically available for earlier papers. The AMS database also contains information on the number of citations received by the papers, but the AMS citation data is incomplete. In particular, it only counts citations in a limited number of journals (which include the most important journals in mathematics), and only reports the post-2000 citations received by a paper (regardless of when the paper was published).

The AMS database has two features that make it invaluable for the type of empirical analysis that we conduct in this paper. First, the AMS devotes significant resources to ensure that every person who has ever published in mathematics is assigned a unique numerical ID that allows him to be distinctly identified even from other mathematicians who share the same name.[19] Second, the editors of *Mathematical Reviews* assign each publication in mathematics to one of the many fields that make up the study of mathematics. The AMS provided us with the author-year-field information at the two-digit field level, identifying publications by each mathematician in each of 63 different fields. In short, the AMS data gives us a complete history of a mathematician's publications, affiliations (since around 1984), and fields.

Our second data source is the Thomson Reuters' Institute for Scientific Information (ISI) Web of Science archive. This archive records the titles, publication source, author

---

[19] As an example of this precision, we learned from personal communication with Professor Victor Kac of MIT that after he defected from the Soviet Union in 1977, he had to publish his work with his Soviet advisor (who remained in the USSR) under the Italian pseudonyms Gatti and Viniberghi, since otherwise the paper would need special permission from the authorities to be published abroad. Despite the difference in names between Gatti and Kac, the AMS database correctly lists the article under the unique author identifier for Victor Kac.

names, references, and citations of millions of articles from thousands of journals

worldwide. For many articles, the database reports research and reprint addresses for each

author (and the location information is systematically available since around 1978). Most

important, the ISI database contains complete citation information for each article in a

primary set of over 7,000 journals, selected to include all of the most important journals in

each field. Articles in marginal journals also appear in the database if they either cite an

article in the primary database or are cited by such an article.

We purchased the records of all 1.2 million articles in the primary ISI Web of

Science database between 1970 and 2009 for the following categories: Mathematics,

Applied Mathematics, Interdisciplinary Applications of Mathematics, Mathematical Physics,

and Statistics and Probability. We also purchased the records of the additional 4.4 million

articles that are either referenced by these main articles or that cite these main articles.

The American Mathematical Society then gave us permission to conduct a paper-by-paper

match of our ISI database with the AMS internal archives. We obtained 882,088 matches

out of the 1,753,148 journal articles in the AMS database for the relevant period, or just

slightly over a 50 percent match rate.

The AMS/ISI databases provide two alternative measures of citation counts for

mathematicians. Neither measure is perfect. The AMS citation data is limited in scope; it

canvasses only a select number of journals and only includes the post-2000 citations made

to mathematical papers. The ISI citation data has a much broader scope (many more

journals) and counts all citations to a paper (not just those made after 2000). However, ISI

citation counts are the result of an imperfect matching process that may be missing a

sizable number of mathematicians and papers.[20] In fact, the matching process is particularly problematic for foreign countries where the spelling of authors' names creates matching difficulties (in particular, the Soviet Union). Not surprisingly, the match rate for papers written between 1984 and 1989 by an author using an American affiliation is 59 percent, while the respective match rate for papers written for authors using a Soviet affiliation is 16 percent. Our empirical evidence is robust to the citation measure we use. Nevertheless, we will generally use the ISI citations data when we examine the productivity of American mathematicians, and use the AMS data when we make international comparisons.

Finally, the Mathematics Genealogy Project (MGP) gave us access to their entire archives. The MGP provides detailed information on doctoral degrees awarded in mathematics since the 14th century. The record for each degree contains the name of the mathematician, the name of the advisor, the title of the dissertation, and the name and location of the institution granting the doctoral degree.

The MGP data, of course, allow a detailed construction of the genealogical tree in the development of mathematics since it can completely link any given mathematician to all his intellectual ancestors and descendants. Equally important, the AMS and MGP have worked jointly to ensure that the unique author ID developed by the AMS can identify mathematicians in the MGP database. By using this unique identifier, we then matched

---

[20] There is a steady increase in the match rate over time, from about 40 percent for journal articles published in the late 1970s to around 75 percent for articles published after 2002. We are able to match over 90 percent of the papers for the most prolific mathematicians during the period, suggesting that some of the missing articles are in marginal journals not covered by the ISI database. As a result, the citation-weighted count of the number of missing articles is likely to be low.

every single advisor and student in the MGP data to their publication histories in the AMS data (and to their citations history in the ISI).

We use these data to construct samples of "active" American mathematicians and Soviet émigrés. The population of mathematicians who were active prior to the collapse of the Soviet Union is given by the sample of persons who published at least one paper between 1970 and 1989.[21] We use the AMS or ISI data to establish the location of the affiliations associated with the various publications in the mathematician's history. A "predominantly" American mathematician is someone who used an American affiliation more than half the time in the pre-1990 period.[22] If the AMS/ISI data do not contain any usable information on the mathematician's affiliations, we then use the MGP data to determine the country where the mathematician received his or her doctoral degree. Prior to 1978, over 85 percent of mathematicians who received their doctoral degrees from an American institution were either American citizens or permanent residents. Moreover, 65.3 percent of the small group of doctoral recipients who were temporary residents stated that they intended to stay in the United States.[23] Hence we supplement the sample of predominantly American mathematicians by also including persons who, in the absence of any specific affiliation information between 1978 and 1989, received their doctoral degree from an American institution between 1960 and 1978.

---

[21] In addition, we restrict the sample to mathematicians whose first publication appeared in print in or after 1940.

[22] More precisely, the mathematician used an American affiliation in more than half of the papers published prior to 1990.

[23] These statistics are calculated using the NSF Survey of Earned Doctorates, an annual survey that contains demographic information for the recipient of every doctoral degree awarded in the United States since 1957.

The definition of the universe of Soviet mathematicians follows analogously. We first use the AMS or ISI data to determine if the mathematician reports a Soviet affiliation. We classify anyone who used a Soviet affiliation more than half the time prior to 1989 as a "Soviet mathematician." In the absence of any affiliation data, we again supplement the sample by including all the mathematicians who received their doctoral degree from a Soviet institution between 1960 and 1989. From this universe, we then define the subsample of Soviet émigrés as the group whose modal affiliation after 1992 was in a country outside the geographic boundaries of the former Soviet Union.[24]

These definitions yield a pre-existing population of 29,392 predominantly American mathematicians and 12,224 Soviet mathematicians. We also estimate that 1,051 of these Soviet mathematicians emigrated after the collapse of the Soviet Union, with 336 of the émigrés moving to the United States.

Figure Va illustrates the frequency distribution of publications for the sample of predominantly American mathematicians, conditional on the mathematician having published at least one article between 1978 and 2008. Not surprisingly, the modal number of publications is only one paper for the entire period. Similarly, Figure Vb shows that the modal number of citations for mathematicians who have published at least one paper is zero. The modal outcome for active mathematicians, therefore, is to publish one paper over a long career span, and for that paper to go unread or, at least, un-cited.

Table I reports summary statistics for the samples of predominantly American mathematicians, Soviet mathematicians who emigrated to the United States, Soviet

---

[24] By construction, an émigré must have published after 1992 (otherwise, we would never observe his post-1992 affiliation).

mathematicians who emigrated to other countries, and the residual group of Soviets who

remained within the geographic boundaries of the former Soviet Union. The data reported

in Table 1 can be used to contrast the productivity of Soviet émigrés in the United States to

that of other Soviet mathematicians and to that of pre-existing American mathematicians.

The most striking finding is the very large degree of positive selection that characterizes

the group of Soviet émigrés in the United States.

Prior to their migration to the United States, the future émigrés were significantly

more productive in terms of the number of papers published and citations received. For

example, before 1992 the average (future) émigré had published 10 more papers and

received 66 more citations (in the AMS data) than the typical Soviet who remained in the

Soviet Union. He had also published 3 more papers than the Soviets who emigrated

elsewhere, and had received over 40 more citations. After 1992, the émigrés' productivity

in the United States far surpassed that of the pre-existing American mathematicians.

Between 1992 and 2008, the average Soviet émigré published 20 more papers than the

average American, and those papers received 143 more citations. In short, the Soviet

émigrés originated in the upper tail of the skill distribution of mathematicians in the Soviet

Union and quickly moved into the upper tail of the skill distribution in the American

mathematics community.

Table I also documents that the Soviet émigrés differ in other dimensions. Consider,

for example, the breadth of a mathematician's interests, as defined by the number of fields

in which the mathematician has published over his or her career. The median number of

fields of publication for both the American mathematicians and the Soviet mathematicians

who remained in the Soviet Union is 2.0. In contrast, the median number of fields for Soviet

émigrés in the United States is 5.5. It is very unusual for a mathematician to publish in more than 5 distinct fields, and only a small (and select) group of American mathematicians do it.[25]

The Soviet émigrés in the United States are also somewhat older. Nearly 40 percent of the émigrés received their degree after 1980. In contrast, 45 to 50 percent of American and other Soviet mathematicians received their degree after 1980.

Finally, Figure VI illustrates how the entry of the émigrés into the American mathematics community affected the supply of pre-existing mathematicians in different fields differently. The "supply shock" is calculated separately for two different types of fields: Soviet-style and American-style fields. Soviet-style fields are the 10 most relatively popular fields in the Soviet Union (i.e., the 10 fields at the top of Figure 1), while American-style fields are the 10 most relatively popular fields in the United States (i.e., the 10 fields at the bottom of Figure 1). We then estimated the output share of Soviet papers in each of these categories by year.[26]

Figure VI shows that the supply shock, as defined by the fraction of total papers published by Soviet émigrés, was very large for the fields emphasized by pre-shock Soviet mathematicians, and was very small for the fields that were dominated by American

---

[25] American mathematicians who publish in many fields are far more prolific than those who specialize in a small number of fields. On average, mathematicians who published in only one or two fields published a total of 6.1 papers between 1978 and 2008. The average total number of papers for mathematicians who published in at least 10 fields is 92.2.

[26] Specifically, the denominator of each fraction is the number of papers in the given subfields published in the United States by authors who were active before 1990, but never had a Soviet affiliation. The numerator of each fraction is the number of papers in the given subfield published in the United States by authors who were active before 1990 and had a former Soviet affiliation. A small group of "defector" Soviets arrived before 1990, mostly in the late 1970s to early 1980s. This group was only about one-tenth the size of the group of émigrés who arrived after the collapse of the Soviet Union. Figure 6 includes the paper-weighted impact of these defectors during their time in the United States, confirming that the paper-weighted size of the influx of Soviet émigrés after the fall of communism was far greater than that of the prior arrivals of defectors.

mathematicians. The fraction of papers published in the United States by Soviet émigrés in "Soviet-style" fields rose from a negligible fraction before 1992 to about 12 percent. In contrast, the fraction of papers published by the Soviet émigrés in "American-style" fields was below 4.0 percent throughout the post-1992 period. In sum, not all American mathematicians were equally affected by the Soviet influx, and it is this differential shock that provides the strategy for identifying the productivity impact of the collapse of the Soviet Union on American mathematics.

IV. Measuring the Productivity Impact

We now examine the extent to which "field overlap" between Soviet and American mathematicians altered the productivity of the latter after the émigrés arrived.

There are many channels through which such productivity effects can arise. The flood of new ideas and theorems could perhaps have spawned a new "Golden Age" as the pre-existing American mathematicians digested and incorporated the new knowledge into their research agendas. At the same time, however, the total number of mathematics faculty jobs, as well as the fraction of resources that deans and administrators allocate among the various subfields of mathematics, is likely constrained and relatively inelastic. The sudden presence of experienced and highly productive Soviet émigrés who may compete for jobs with newly minted doctorates would then almost certainly have a "crowd-out" effect on the paid research jobs that American mathematicians would have otherwise filled.[27]

---

[27] There are other sources of scarcity that can create crowd-out effects, such as the limited attention span of field leaders. There are insurmountable constraints on how much new knowledge a scientist can absorb, so that the location of the "marginal" article that gets ignored moves upwards in the distribution of publications after the influx of Soviet mathematicians. Reasonably good articles that would otherwise have generated some attention will then remain unread and un-cited.

The identification strategy is built on comparing different groups of American mathematicians with each other over time: those who had worked on problems where the Soviets could offer assistance and/or competition, versus those who had worked on problems that the Soviets knew little about. To quantify the degree of exposure, we calculate an index reflecting the field "overlap" between the pre-1990 publication record of *each* American mathematician and that of the Soviets.

We use 3 alternative indices to demonstrate the robustness of our empirical findings. The first index is simply a correlation coefficient calculated for each American mathematician in our data. Let a_{ij} be the share of papers that mathematician *i* published in field *j*, and let s_j be the share of all Soviet papers published in field *j* before the collapse of the Soviet Union.[28] Our first index is simply the correlation coefficient ρ_i between these two vectors.

A second index measures the "intensity" of the overlapping research interests. In particular, let S_j be the total number of papers written in field *j* in the Soviet Union. If we calculate the dot product between the vector S and the vector of shares a_{ij} for each American mathematician, the resulting number gives the "effective" number of Soviet papers that the typical American mathematician would have written, weighted by Soviet preference. To construct an index that lies between zero and one, we divide the dot product

---

[28] The share a_{ij} is calculated using all papers published by a particular mathematician between 1960 and 1989; this calculation uses the database the AMS provided to us. The Soviet share s_j is obtained directly from *MathSciNet* (an online database maintained by the American Mathematical Society) and contains papers published mainly between 1984 and 1989. As noted above, the AMS began to collect systematic information on the country of publication only around 1984 (although a few pre-1984 papers report the affiliation of the author).

by the number of papers published in the most populated Soviet field. Hence the "index of intensity" is given by:

$$(1) \qquad I_i = \frac{\sum_j a_{ij} S_j}{\max(S_j)}.$$

Note that this index will be equal to zero when the American mathematician publishes in fields where Soviets have never published and is equal to one when the American mathematician publishes exclusively in the field where the Soviets have done the most work.

Finally, we employ the commonly-used "index of similarity" (Cutler and Glaeser 2007) defined by:

$$(2) \qquad D_i = 1 - \frac{1}{2} \sum_j |a_{ij} - s_{ij}|.$$

The index of similarity equals one when there is a perfect overlap in the relative field distributions between American mathematician i and the Soviet research program, and zero when there is total dissimilarity.[29]

Table II reports the value of the overlap indices for the most productive American mathematicians, with the ranking determined by the number of papers published between

---

[29] If the complement of the index of similarity (or $1 - D$) were calculated for the pooled group of American mathematicians, its value would give the fraction of American mathematicians who must move across fields in order to ensure that American mathematicians have the same field distribution as the Soviets.

1960 and 1989. There is a great deal of dispersion in the various indices even among these superstars. The correlation coefficient, for instance, ranges from zero to .6, while the index of similarity ranges from .07 to .41.

To estimate the net impact of the Soviet influx on the marginal product, $y_i(t)$, of American mathematician i in year t, consider the regression model:

$$(3) \qquad y_{it} = \phi_i + \phi_t + X_i(t)\, \gamma + \theta\, (T \times Index_i) + \varepsilon_i(t),$$

where ϕ_i is a vector of individual fixed effects; ϕ_t is a vector of year fixed effects; X is a vector of standardizing variables that include the mathematician's years of work experience introduced as a quartic polynomial; T is a dummy variable indicating if the year for the particular observation is 1992 or beyond; and $Index$ is one of the three alternative overlap indices defined earlier.[30] The standard errors are clustered at the individual level.

We use two alternative dependent variables in the analysis: the number of papers mathematician i published in a particular year, and the total number of ISI citations received by the papers published in that year.[31] It is worth noting that the citations data is truncated because a paper may continue to be cited into the future (and some important

---

[30] Of course, the regression does not include either the value of the overlap index or the post-1992 indicator variable. The individual fixed effects subsume the person-specific overlap index, and the period fixed effects subsume the post-1992 indicator. The mathematician's years of work experience is defined as years elapsed since the mathematician's first publication. The regressions only include those observations where a mathematician has at most 60 years of potential work experience.

[31] Our treatment of coauthored papers is as follows: Regardless of how many authors are responsible for a single paper, each author is given full credit when we count a mathematician's number of papers and citations. This approach is required by the nature of the AMS data. In particular, we have counts of papers by year, field, and author, and these counts do not allow us to identify all the authors of any specific paper. The only exception to this method of treating coauthorships is the construction of Figure I (and the baseline field distribution of Soviet mathematical research used in the calculation of the overlap indices), where we obtained the total number of papers by field directly from *MathSciNet*.

papers may not have been recognized by the end of our sample period).[32] Note, however,

that the inclusion of year fixed effects in equation (3) helps to control for this problem.

The identification strategy used by the regression model in (3) can be easily

described. We are examining how the research output of mathematician i changed after

1992, when Soviet and American mathematicians began to interact both physically *and*

intellectually. The coefficient θ measures whether the marginal product of mathematicians

who had pursued a Soviet-style research agenda changed after 1992, when the Soviet ideas

began to be disseminated to a wider audience in the United States, and Soviet

mathematicians began to publish their ideas in the same journals and compete for jobs in

the same institutions as American mathematicians. If the externalities arising from the

entry of the highly skilled Soviets specializing in a relatively small set of mathematical

fields are weak or non-existent, we would expect that the American "intellectual

competitors" of the Soviets would become less productive as a result of the influx and the

coefficient θ would be negative.

The control group in our empirical exercise consists of American mathematicians

who prior to 1990 did not specialize in the fields dominated by the Soviets. This control

group not only experienced a smaller Soviet influx post-1992 (see Figure VI), but it also did

not gain directly from the influx of new ideas. Specifically, the exposure to Soviet ideas was

unlikely to be useful to American mathematicians working in U.S. dominated fields: out of

all the references cited by articles in the top 10 U.S. dominated fields, only 5 percent of

these references are to articles in Soviet-dominated fields. Therefore, there is little

---

[32] The total number of citations received by a paper published in year t is measured as of calendar year 2009.

evidence of complementarities across the intellectually distant fields pursued by Soviet and American mathematicians.

Table III reports the estimates of the coefficient θ obtained from a variety of different regression specifications. In addition to the two dependent variables and three measures of the overlap index, we estimated the regression on two alternative samples of American mathematicians: the sample of predominantly American mathematicians defined earlier, as well as a sample of "exclusively" American mathematicians. The sample of exclusively American mathematicians is the subset of predominantly American mathematicians who were always affiliated with an American institution before 1990.[33] It is important to note (and is obvious from biographies of the superstars listed in Table II) that either definition of an "American" mathematician includes large numbers of foreign-born persons. These foreign-born mathematicians may have either migrated to the United States after publishing some initial work abroad, or first arrived in the United States as foreign students and stayed in the country after receiving the doctorate.

Regardless of the various specification changes, the results reported in panel A of Table III are robust. In particular, all of the overlap indices have a strong negative impact on the post-1992 productivity of American mathematicians. It is instructive to give a numerical example to emphasize that the quantitative impact on the number of publications is numerically important. In terms of the correlation coefficient, for instance, an increase from $\rho = 0$ to $\rho = 1$ (so that the pre-1990 field distribution shifts from one where the American mathematician's work is uncorrelated with that of the Soviets to one where the mathematician's research interests perfectly overlap) reduces the number of

---

[33] There are 27,663 persons in the sample of exclusively American mathematicians.

publications in any particular year by 0.13. The regression, therefore, predicts that the

average American's productivity fell by 2.2 papers (or 0.13×17) over the 1992-2008

period. As Table I shows, the average American mathematician published 6.8 papers during

the period, so that the Soviet supply shock reduced productivity by around a third.

It is easy to demonstrate graphically the impact of the supply shock on the

productivity of American mathematicians with overlapping research agendas. Suppose we

classify mathematicians whose index of similarity is in the upper quartile of the

distribution as "highly exposed," while the group in the bottom quartile has low exposure.

As shown in Figure VIIa, the raw trends in the average number of papers published by the

average mathematician in each of these two groups are revealing. Prior to 1990, the highly

exposed group had a slight upward trend, while the least exposed group had a slight

downward trend. After 1990, however, there is a precipitous (absolute and relative)

decline in the publication rate of the group whose research agenda overlaps most with the

Soviets. The trends in the adjusted number of papers (after removing individual fixed

effects) shown in Figure VIIb reveal a continuation of previous trends for the least exposed

mathematicians, and a striking reversal of trends for the highly exposed mathematicians.

Table III also reports the estimates of the coefficient θ when the dependent variable

is the ISI number of citations (or roughly the number of publications weighted by

quality).[34] The evidence documents a large decline in the number of citations written by

American mathematicians whose research interests greatly overlap with those of the

Soviets. In other words, not only are the competing mathematicians producing fewer

---

[34] We also estimated the regressions using the AMS count of citations as the dependent variable. As with the ISI data, all of the interaction coefficients are negative and significant regardless of the measure of overlap used.

papers, but also the work they produce is of "lower quality" in the sense that it generates many fewer citations. A one-unit shift in the correlation coefficient would reduce the number by about 20 citations per year.

The next two panels of Table III address the question of whether these effects persist in the long run. In panel B, we restrict the analysis to the years 1978-1999, so that the regressions measure the short-run productivity effect, while panel C only includes the years 1978-1989 and 2000-2008, so that the productivity impact is measured roughly 10 to 15 years after the Soviet influx occurred. The perhaps surprising finding in the table is that, if anything, the long run effects are larger than the short-run effects (although the difference is sometimes not statistically significant). It is typically very difficult for academics to reenter the publications market once they have taken some years off from successful active research. In academia, the short run is the long run.[35]

A potential problem with these regression results is that although they include year fixed effects to net out any trends in total output in the mathematics discipline, there may be field-specific time trends that may be spuriously correlated with the Soviet influx. For instance, Soviet-style fields may have experienced a differential response in funding or intellectual attention after the cold war (though we have found no evidence for this in our historical research). To determine if the results are sensitive to the inclusion of arbitrary

---

[35] In the long run, we would have expected resources to be moved to the affected fields. It is unclear, however, that such adjustments took place. In fact, federal obligations to universities and colleges for research in mathematics declined slightly (in real dollars) during the 1990s; see U.S. National Science Foundation (2004, Table 1a). Anecdotal evidence also suggests that it was difficult to obtain additional resources for hiring Soviet mathematicians. Personal communication with Professor Arthur Jaffe provides two such examples. Professor Jaffe was Chair of the Harvard mathematics department at the time and requested resources to hire two Soviet mathematicians classified as "Targets of Opportunity." The Dean of the College denied the request, although Harvard President Derek Bok belatedly reversed that decision. Similarly, Jaffe and Bok contacted 54 foundations requesting financial assistance to fund the transition of Soviet mathematicians into the American mathematics community. All of the funding requests were denied.

field-specific trends, we re-estimated the model with a full set of field-year fixed effects in a sample constructed so that a particular observation represents an author-field-year cell. In other words, our data now consists of the publication history of each American mathematician in each of the 63 fields in mathematics for each year in our sample period. Of course, many of the values in the dependent variable in this regression will be zero simply because few mathematicians publish anything at all in a particular year, and fewer still publish a paper in more than one field.

The generalized regression model is given by:

(4) $\qquad y_{ij}(t) = \phi_i + \phi_j + \phi_t + (\phi_j \times \phi_t) + X_i(t)\,\gamma + \theta\,(T \times Index_i) + \varepsilon_{ij}(t),$

where $y_{ij}(t)$ is a measure of the marginal product of mathematician i in field j at time t; ϕ_j is a vector of field fixed effects; and $(\phi_j \times \phi_t)$ represents all possible interactions between the field and year fixed effects.

The bottom panel of Table III summarizes the regression results. It is evident that the coefficient θ remains negative and significant even after controlling for the field-year fixed effects, regardless of the overlap index used. The numerical magnitude of the coefficient is smaller than that of the analogous coefficient reported in Panel A, but this is simply a mechanical effect because the total impact on publications is divided among the 63 fields that make up the mathematics discipline. Once we scale the coefficient properly (by multiplying by 63), the regression coefficients reported in Panel D are, in fact, remarkably similar to those in Panel A. The regression results, therefore, indicate that arbitrarily complicated field-specific trends do not account for the post-1992 break in the

evolution of publications for mathematicians whose research interests greatly overlapped with that of the Soviet émigrés.

Although Table III describes how the collapse of the Soviet Union affected the mean productivity of American mathematicians, it is of interest to measure the impact at extreme points in the quality distribution of the pre-existing workforce. Consider initially the impact of the Soviet influx at the top of the quality distribution. Although the mean American mathematician was negatively affected, those losses could perhaps be more than offset by an increase in the likelihood that top American mathematicians produced more "home runs." We define a home run as a paper that is in the top tail of the distribution of citations among *all* papers published in a particular year.[36] To assess the sensitivity of the results, we use three alternative definitions of the number of home runs: the number of papers that the mathematician has published in any given year that lie above the 90th, the 95th, or the 99th lifetime citation percentile for papers published that year. Since the signal value of citations received by recent publications is very weak (i.e., the papers have not been out long enough to be cited), we restrict the analysis to papers published between 1978 and 2005, a period where we can reasonably distinguish between home runs and run-of-the-mill papers.

By construction, a home run is a very rare event. After all, the modal number of publications for an active mathematician in any given year is zero. As is common in the literature that examines these types of rare publication events, we use a Poisson regression

---

[36] To estimate the number of home runs published by mathematician *i* in year *t*, we first calculated the total number of (worldwide) lifetime citations received by each article in the ISI database that had been successfully matched to the AMS database. We used these lifetime citation totals to calculate the 90th percentile, 95th percentile, and 99th percentile thresholds of the citations distribution for papers published in each year. We then used the matched AMS/ISI database to count the number of papers that a particular mathematician published each year that exceed these threshold percentiles.

model to examine how the collapse of the Soviet Union affected the number of home runs published by American mathematicians. The Poisson model can be interpreted as a regression where the dependent variable is the expected value of the number of rare events attained by a particular observation. In particular, let:

(5) $$E[\log H_{it}] = \phi_i + X_{it}\,\beta + \delta\,T + \varepsilon_{it},$$

where H_{it} is the number of home runs achieved by mathematician i in year t; ϕ_i is an individual-specific fixed effect; and X_{it} is a vector of covariates, again including the mathematician's years of work experience (introduced as a quartic polynomial). To clearly illustrate the source of the additional (or fewer) home runs hit by American mathematicians after the collapse of the Soviet Union, we estimated the Poisson regression separately for three different groups: the group of least exposed mathematicians (below the 25th percentile in the distribution of the respective measure of the index of overlap); the group of most exposed mathematicians (above the 75th percentile), and the residual "middle" group. The coefficient of interest is the parameter δ, which measures the post-1992 break in the home runs trend for the various groups of American mathematicians.

The first three columns of Table IV report the coefficient of the post-1992 indicator variable for the various specifications. It is evident that the expected number of home runs written by mathematicians who most overlapped with the Soviets declined dramatically after 1992. For example, consider the coefficients estimated when the mathematicians are classified according to their placement in the distribution of the correlation coefficient. If we define a home run as a paper above the 95th percentile of citations, the coefficient for

the most exposed group is -0.204 (0.032), indicating that a mathematician in this group experienced a 20.4 percent decline in the expected number of annual home runs hit after 1992. In contrast, the coefficient for the least exposed group is 0.058 (0.066), suggesting that the least exposed mathematicians had just about as many home runs before and after the collapse of the Soviet Union. In fact, all the regression models predict an absolute (and numerically large) decline in the number of home runs published by the most exposed mathematicians. In contrast, the least exposed mathematicians usually did not experience a change in the number of home runs (although the positive coefficient is significant when using the index of similarity).

Of course, it is possible that the "average" adverse impact of the Soviet influx on the number of home runs masks a potentially important gain within a small subgroup of American mathematicians. Elite institutions hired many of the émigrés. These émigrés began to coauthor with their American colleagues (as suggested by Figure II). In fact, 666 pre-existing American mathematicians (or 2.8 percent) coauthored with a Soviet émigré at some point after 1992.[37] However, the coauthorship rate was 5.7 percent among highly exposed American mathematicians, and rose further to 8.0 percent among the highly exposed mathematicians affiliated with the top 25 institutions.

To determine if the high rate of coauthorship among highly exposed American mathematicians led to an increase in the number of home runs hit by these Americans, we re-estimated the regression in equation (5) separately by coauthorship status. The

---

[37] The incidence of coauthorship can only be estimated in the matched AMS/ISI database. A highly exposed mathematician is someone who is in the top quartile of the distribution of the index of similarity. The institutions were ranked according to the total number of papers published during between 1984 and 2008 by all mathematicians affiliated with the institution.

estimated coefficients are reported in the last two columns of Table IV.[38] There is an important difference in the impact between the Americans who coauthored with the émigrés and those who did not. In particular, the evidence clearly shows that coauthoring with a Soviet émigré entirely attenuates the negative impact of the Soviet influx on the number of home runs written by highly exposed American mathematicians. Nevertheless, there is no evidence of any increase in the number of home runs hit by American mathematicians even within the small subset of highly exposed coauthors.

It is equally important to measure the productivity impact at the other end of the quality distribution, and examine what happened to the "marginally active" American mathematician. In particular, we examine the probability that an American mathematician remains active (i.e., publishing) in the profession after 1992. We constructed a variable indicating if the mathematician has "retired" by setting it equal to zero in every year prior to the last year in which we observe a publication for mathematician i. We then omit all post-retirement observations for the particular mathematician (in other words, we exclude the years after the publication of a mathematician's last article). We used this form of survival data to estimate Cox proportional models where the dependent variable is the instantaneous probability of retirement:

(6) $\qquad \log \lambda_i(Z_i, \tau) = \log \lambda_0(\tau) + Z_i \beta + \varepsilon,$

---

[38] Although the coauthorship evidence reported in Table IV uses the 95th percentile threshold to define a home run, the qualitative results are similar for the other thresholds. Note that coauthorship status is not randomly assigned across mathematicians, so that the coefficients, though suggestive, do not necessarily prove a causal relationship.

where $\lambda_i(Z_i, \tau)$ is the hazard for individual $\underline{i}$ with covariates Z_i at experience τ; and $\lambda_0(\tau)$ is

the baseline hazard for workers at that level of experience. The covariates in the vector Z

include the mathematician's overlap index, the post-1992 indicator variable, and the

interaction between these two variables. The top row of Table V reports the coefficient of

this interaction variable estimated separately for the three alternative measures of the

overlap index.

The overlap index has a positive and significant impact on the instantaneous

retirement probability after 1992, and the implied numerical effect is large. Consider, for

example, the coefficient estimated when we use the correlation coefficient. The regressions

imply that a mathematician whose prior research was perfectly correlated with that of the

Soviets had roughly a 50 percent (given by $e^{+0.41}$) increase in the instantaneous probability

of retirement relative to the probability of a mathematician whose work was uncorrelated

with the Soviet research program. It is easier to get a sense of the quantitative effect by

examining the survival functions implied by the estimated Cox proportional hazards.

Figures VIIIa and VIIIb illustrate the survival functions for high- and low-exposure

mathematicians, both before and after 1992.[39] The probability of survival declined

precipitously for high-exposure mathematicians after 1992 (both relatively and

absolutely).

The shape of the survival functions, in fact, suggests that the effect seems largest for

mathematicians at early points in their career. We re-estimated the hazard regressions

separately on mathematicians classified by years of experience. The remaining rows of

---

[39] The low- and high-exposure groups are defined by the bottom and top quartiles of the distribution of the index of similarity.

Table V show that the adverse impact of an overlapping research agenda with the Soviets is far larger in the sample of younger (and likely untenured) mathematicians. In other words, the Soviet supply shock had a particularly detrimental effect on the long-run career prospects of young (and untenured) mathematicians, greatly increasing their probability of retiring from the production of mathematical output.[40]

Although some American mathematicians stopped publishing altogether after the collapse of the Soviet Union, other mathematicians remained active but were "encouraged" to move to different institutions. We create a variable for each mathematician indicating if the modal institution in the post-1992 period (after the Soviet influx) was different from the last institutional affiliation observed prior to 1990.[41] By construction, our sample consists of mathematicians whose last reported affiliation prior to 1990 and whose modal affiliation post-1992 were located in the United States. We then estimated linear probability models that examine the separation probability of a mathematician, including controls for differences in individual and institutional quality. In particular, we hold constant the log number of papers the mathematician published prior to 1990, the log number of papers published by mathematicians affiliated with the pre-1990 institution, and the mathematician's years of work experience.

The regression model uses two variables to capture the impact of the Soviet influx. The first is the overlap index that describes the pre-1990 overlap in research interests between the particular American mathematician and the Soviet research program. The

---

[40] The regression obviously cannot distinguish between the possibility that the mathematician stopped writing publishable articles or was simply unable to find an outlet for his research in the face of increasing competition by the highly productive Soviet mathematicians.

[41] The construction of the sample requires that the mathematician published both before and after 1992, and that there is a valid affiliation code in both periods.

second is a dummy variable indicating if the institution employing the mathematician just before the shock was, in fact, *subsequently* "exposed" to Soviet émigrés (i.e., if the institution hired a Soviet émigré at some point after 1992).

The top panel of Table VI reports the regression coefficients.[42] The separation rate is strongly related to both geographic and intellectual proximity. In particular, the separation rate rises by about 5 percentage points if the mathematician's last affiliation before 1990 ended up hiring a Soviet émigré.[43] Similarly, regardless of how the overlap index is measured, there is a strong positive correlation between the index and the separation rate. An increase in the correlation coefficient from zero to one increases the separation rate by 17 percentage points.[44] Put differently, holding geographic exposure constant, the American mathematicians with the most Soviet-like research agendas suddenly found themselves in volatile jobs and ended up moving to other institutions at dramatically higher rates.

The bottom two panels of the table replicate the mobility analysis for the samples of "untenured" and "tenured" mathematicians. Notably, the separation rate of younger faculty

---

[42] Although we do not report the coefficients of the other variables in the regression, they mirror findings in the labor economics literature. The higher the quality of the mathematician (as measured by the number of pre-1990 publications), the lower the separation rate. This finding is analogous to the negative correlation between the probability of job separation and a worker's wage. In addition, the separation rate is lower for older mathematicians and higher for mathematicians initially employed by high-quality institutions.

[43] There may be reverse causality in the correlation between the separation rate and whether the institution eventually hired a Soviet émigré. Perhaps the institution only hired émigrés after an American mathematician failed to make tenure. Although this interpretation can explain the correlation, it is important to note that the failed assistant professors were young and inexperienced, while the Soviet émigrés were more senior and many had already had distinguished careers.

[44] We also estimated separation regressions that included a vector of fixed effects for the pre-1990 institution. Although the impact of geographic exposure can no longer be identified, the coefficient of the overlap index remains negative and significant. In the regression that uses the correlation coefficient, for instance, the coefficient is 0.123 (0.025), so that the degree of research overlap with the incoming Soviets determined which mathematician within a particular institution was most likely to leave.

is much more sensitive to direct exposure to the Soviet influx. In particular, the probability of moving increases by about 10 percent for a young mathematician who was employed in an institution that eventually hired a Soviet émigré, but is unaffected for an older mathematician. Similarly, the impact of the overlap index is at least twice as large in the untenured sample, regardless of how the index is measured.

The mobility regressions raise a number of interesting questions. Most important: Where did the displaced mathematicians go? We calculated a measure of change in institutional quality by adding the total number of papers published by all mathematicians affiliated with a particular institution throughout the entire 1984-2008 period. Since we can identify the last institution employing the mathematician prior to the supply shock as well as the modal institution employing him after the influx, we can construct a measure of the change in institutional quality for movers. Let N_0 be the number of papers published by the institution employing the mathematician before 1990, and N_1 be the number of papers published by the institution employing him after the influx. Our measure of the change in institutional quality is defined by $\log(N_1/N_0)$.

Table 6 shows that this change in institutional quality is related to both geographic and intellectual proximity, with both variables having strong negative effects. The impact of geographic proximity is particularly large. Conditional on moving, an American mathematician initially employed by one of the institutions that ended up hiring Soviet émigrés moved to an institution that had published 91 percent fewer papers during the period.[45]

---

[45] Although the numerical magnitude of this effect seems implausible, it is consistent with the raw data. The total mathematical output produced by institutions falls precipitously once we move from the top-ranked

Finally, up to this point our analysis has used papers and citations to measure a mathematician's marginal product. Academic mathematicians also produce students, and the collapse of the Soviet Union may have had beneficial or harmful effects on the quality of the students produced by American institutions. Most obviously, the hiring of highly skilled Soviet mathematicians should presumably lead to higher quality students. At the same time, the Soviet influx had an adverse effect on the productivity of mathematicians who had previously done Soviet-style research, and this adverse impact could be transmitted "intergenerationally" to their students.[46]

Despite the endogenous sorting of advisors and students, it is of interest to examine how the quality of the students was affected by the "Soviet-ness" of the advisor: either because the advisor was himself a Soviet émigré or because the advisor was an American doing Soviet-style research.[47] As noted above, the MGP data links each new mathematician to his or her advisor. The AMS/ISI data further allows us to calculate the average annual number of papers and citations produced by each student between the year the degree was awarded and 2011.

The data consists of a single observation for each student obtaining the doctoral degree from an American institution between 1978 and 2008, and who had either a Soviet or a predominantly American advisor. We estimated a regression model linking the student's (future) annual productivity to variables characterizing the Soviet-ness of the

---

institutions to even those ranked 200th in our index of quality. Mathematicians affiliated with the top institutions (e.g., MIT or the University of California, Berkeley) published over 8,000 papers during the period. Mathematicians associated with institutions around the 200th rank published only around 350 papers.

[46] One could also argue that the most affected American mathematicians (whose publishable product declined after 1992) had more time for student advising and could perhaps produce higher quality students.

[47] The sorting of advisors and students is not random. Great mathematicians will likely attract great students, and/or great mathematicians can be more selective in choosing students from the pool of potential advisees.

advisor: a dummy variable indicating if the advisor is a Soviet émigré, the value of the

overlap index if the advisor is a predominantly American mathematician, and an

interaction between the advisor's overlap index and a dummy variable indicating if the

student received the doctoral degree after 1992.[48] This interaction variable captures the

differential impact on student productivity of having an American advisor who was most

affected by the post-1992 Soviet influx.

Column 1 of Table VII reports the coefficients of the Soviet émigré indicator and the

overlap interaction when we use the student's annual number of papers and ISI citations as

the dependent variables. The results are striking. The presence of a Soviet advisor has a

strong positive effect on both papers and citations, leading to about 0.2 to 0.3 more papers

and 0.5 to 1.0 more citations per year (depending on the measure of the overlap index).

At the same time, however, the Soviet-ness of the student's American advisor has a

mirror-image negative impact on the student's productivity. To easily interpret the

numerical magnitude of this effect, suppose again the overlap index is measured by the

correlation coefficient. The student of an American advisor with a correlation coefficient

equal to 1 publishes 0.4 fewer papers and produces 7.0 fewer citations per year than the

student of an American advisor with a correlation coefficient of zero.

The construction of the data allows us to introduce both institution and advisor

fixed effects, so that we can measure productivity effects that net out the influence of

institution and advisor quality. The remaining columns of the table introduce these fixed

effects. The impact of having a Soviet advisor remains strong even after we include

institution fixed effects in the regression. Of course, we cannot estimate the Soviet advisor

---

[48] The regression also includes a set of year fixed effects.

coefficient if the regression also includes advisor fixed effects. Nevertheless, it is worth noting that the American advisor's overlap index often plays a differential role before and after 1992 even after we include advisor fixed effects. The inclusion of the advisor fixed effect allows us to measure how the Soviet-ness of an American advisor affected the productivity of *this particular advisor's* students over time. The interaction coefficient is negative (and significant using two of the measures of the overlap index). In short, the evidence suggests that the students of a specific Soviet-like advisor were directly and adversely affected by the Soviet influx.

It is easy to manipulate the estimated coefficients to get a back-of-the-envelope measure of the impact of the Soviet influx on the average productivity of students produced by American institutions. For example, about 3.0 percent of students had a Soviet advisor after 1992. Those select students were able to generate around +0.007 more papers per year (or 0.03 × +0.23) as a result of the Soviet influx. The other 97 percent of students, however, had a non-Soviet advisor, and the average advisor's correlation coefficient was .09. The predicted decline in the number of papers produced due to the intergenerational transmission of less marketable skills is around -0.038 papers (or 0.97 × 0.09 × -0.43). The net effect, therefore, seems to be negative.

V. Aggregate Effects

We now shift the focus from estimating individual-level productivity effects to calculating aggregate effects. In particular, we wish to determine if the total number of papers published by the cohort of pre-existing American mathematicians increased or decreased as a result of the Soviet influx. Equally important, did the size of the American

"mathematics pie" expand or contract after we take account of the contribution of Soviet émigrés?

Suppose we divide the pre-existing American mathematicians into three categories: those with a pre-1990 research agenda that was highly correlated with that of the Soviets, those with an agenda that was almost uncorrelated, and everyone else. As before, we use the 25th and 75th percentile of the distribution of the index of similarity to calculate the thresholds that define the groups.

For our first prediction exercise, we estimated separate regression models that trace out the mathematician's age-product profile for each group. Specifically, we estimated a regression of the number of papers published by a particular mathematician in a given calendar year on years of experience (introduced as a quartic polynomial) and on a vector of individual fixed effects. Notably, we only used data from 1970 through 1991 to estimate this regression model. We then used the regressions to predict the post-1992 output of each mathematician. To ensure that our predictions are not extrapolating into the far-off future, we conduct this "short-run" prediction exercise for the years 1992-1999.[49] We then added the total number of predicted publications across years and across all mathematicians in each group. The top panel of Table VIII reports the predicted total number of papers (per year) and contrasts these predictions with the actual output of the pre-existing mathematicians.

---

[49] Since the regression will be used to predict outcomes about a decade out of sample, we use a longer span of the pre-1992 data to get a more precisely estimated age-productivity profile for mathematicians in each group. The regressions are then used to predict annual papers published in 1992-1999 for mathematicians who either did not publish at all after 1992 (so that we do not know their location), or whose publication record reveals that they were affiliated with American institutions during the period.

Not surprisingly, the exercise reveals dramatic differences in how well the pre-1992 productivity history predicts the post-1992 output. Consider, for instance, the group of mathematicians in the bottom quartile of the distribution of the correlation coefficient (i.e., those mathematicians whose work least overlapped with that of the Soviets). Based solely on the path of the pre-1992 age-product profile, we would expect these mathematicians to publish 901 papers between 1992 and 1999. In fact, this group published 1,253 papers. Therefore, the regression model under-predicts the number of publications for mathematicians whose work was uncorrelated with that of the Soviets (and the gap between the prediction and the actual output is statistically significant). However, our prediction is much further off the mark for mathematicians who are in the top quartile of the distribution of the index of similarity. Their pre-1992 publications history predicts that they would publish 5,062 papers annually between 1992 and 1999. In fact, they only published 4,015 papers, a "loss" of over 1,000 papers. This difference is highly significant.

If we add up the total number of predicted and actual papers across the three groups, our exercise over-predicts the actual output by about 600 papers (or 6.8 percent), a statistically significant difference. However, Table 8 also reports that Soviet émigrés published 371 papers in the United States annually between 1992 and 1999. Once we take into account the Soviet contribution, our prediction exercise closely tracks what actually happened: The pre-1990 publication history of the pre-existing American workforce would suggest a total of 9,482 papers published annually between 1992 and 1999. There were, in fact, 9,245 publications, and this small difference is not statistically significant.

An alternative prediction exercise would use the data available over the entire 1978-2008 period to re-estimate the basic regression model in equation (3). We can then

use this regression to conduct a counterfactual exercise that predicts the number of papers for each mathematician in each year between 1992 and 2008 if the Soviet influx had never occurred. We model this counterfactual world as one in which the index of similarity would not play a differential role in determining the mathematician's product before and after 1992.[50]

We again aggregated the predicted and actual number of papers for the three groups of mathematicians, and these aggregate predictions are reported in the middle panel of Table VIII. As with the short-run exercise, we tend to slightly under-predict the performance of American mathematicians who did not overlap with the Soviet research agenda. In contrast, we tend to over-predict by a larger amount (529 papers), the output of mathematicians who were most Soviet-like in their research. This prediction exercise suggests that there should have been a total of 7,548 papers published annually in the post-1992 period. In fact, there were 7,831 annual papers after we include the Soviet contribution. The Soviets, therefore, slightly increased the size of the "mathematics pie" (by 284 papers annually), but the increase is not statistically significant.

The crucial lesson from these prediction exercises is that the total product of the pre-existing American mathematics community contracted as a result of the Soviet influx, and that the Soviet contribution to American mathematics "filled in" the gap. Notably, there is no evidence that the Soviets greatly increased the size of the pie.[51] The papers that are

---

[50] Specifically, the prediction uses the model estimated in equation (3) and sets the interaction of the post-1992 dummy variable with the index of similarity equal to zero.

[51] We also checked this prediction by considering whether the predicted decline in the publication of "home runs" suggested by Table IV is made up by the home runs produced by Soviet émigrés in America. That exercise also suggests that some, and possibly all, of the lost American home runs were made up by the Soviets.

missing from the American portfolio are papers that would have been written by the most Soviet-like Americans, but never were.

The nature of the academic market can help us understand how this situation may arise. First of all, the American Mathematical Society's Annual Surveys report surprisingly small changes from year to year in the total number of mathematicians working in U.S. research institutions during our sample period.[52] With the total number of research jobs unresponsive to the opportunities created by the Soviet influx, the entry of experienced Soviet mathematicians had particularly adverse effects on untenured and/or marginal mathematics faculty. As we have shown, some of those faculty moved to lower-ranked institutions and some of them moved out of the academic market altogether. Regardless of their final placement, many of these mathematicians found it increasingly difficult to do the type of research that leads to publishable output, either because of an increased teaching load, fewer networking possibilities, or because the responsibilities of a job as a "quant" in Wall Street limits the kind of effort required to develop publishable material.

Although we do not find any convincing evidence that there was a substantial increase in the aggregate output of either the pre-existing American workforce or of that community combined with the Soviet émigrés, it is still possible that the "quality" of American mathematics improved in a broader sense. Put simply, if journal editors maximize the visibility of their journals, then the quality of papers published in the United States by Soviet émigrés must have been at least as high as that of papers they replaced.

---

[52] For instance, between the 1990-91 and 1994-95 academic years, the total employment of doctoral full-time faculty in mathematics departments of Ph.D.-granting institutions increased from 6,008 to 6,147. The number of such faculty at the 82 institutions with a Conference Board of Associated Research Councils Rank of I or II (i.e., those institutions that produce the bulk of the best research) actually decreased from 3,740 to 3,613. It is also interesting that during this time period total employment in statistics departments also declined slightly, from 1,123 to 1,099.

We examine this implication by predicting the aggregate number of citations that American mathematicians should have produced after 1992, and determining if any drop in the total number of citations is more than made up by citations to papers written by Soviet émigrés after their arrival in the United States.[53] As the bottom panel in Table VIII demonstrates, there was a sizable drop in the total number of citations produced by American mathematicians, but the increased number of citations produced by the Soviet émigrés more than compensated for the drop—although the net increase is not significant.[54] In sum, the evidence allows for a plausible (though weak) case that the American mathematics market succeeded in replacing some proofs that would otherwise have been proven with a somewhat larger number of slightly more interesting theorems.

But it is problematic to conclude that there was a (small) net benefit to American mathematics because the counterfactual is difficult to observe. In mathematics (unlike in much of, say, economics), a result—i.e., a proof—that is true is true for all time. This suggests that there is an opportunity cost to slightly less interesting proofs never being published in order to make way for slightly more interesting ones; the less interesting proof could have been a foundation for a more important proof in the future. A complete evaluation of costs and benefits, therefore, would need to compare the long-run value of the proofs that were, in fact, proven by the Soviet émigrés with the proofs that would have

---

[53] The prediction exercise uses the AMS counts of citations since the ISI severely undercounts citations by Soviet mathematicians.

[54] It may also be the case that mathematical work is now more often cited outside mathematics. For example, experimental physicists may be citing and using newly available Soviet work on quantum theory in their own research. It is certainly the case that American papers outside mathematics cite mathematical papers. There were 256 thousand such citations in 1978, and this number rose to 1.8 million by 2007. However, citations to Soviet mathematical research are quite rare throughout the entire period. These citations accounted for three-hundredths of one percent of the total in 1978 and seven-hundredths of one percent in 2007. It seems implausible; therefore, that Soviet mathematical research could have formed the basis for large improvements in American non-mathematical fields.

been proven by the displaced American mathematicians, many of whom were relatively young and perhaps at their peak mathematical efficiency.

VI. International Differences

The evidence reported in the previous sections documents that American mathematicians whose research agenda most overlapped with that of the Soviets suffered a loss in productivity after the collapse of the Soviet Union. The regressions, however, do not directly link the decline in productivity to the physical presence of Soviet émigrés in the United States.

As in studies of the labor market impact of immigration, it can plausibly be argued that some, if not all, of the impact would have been observed because of the increased "globalization" of the mathematics marketplace after 1992, regardless of how many émigrés the United States actually admitted. After all, the most highly substitutable American mathematicians could have become less productive simply because the Soviets (regardless of where they are located) were now free to compete for journal space, financial grants, and other types of academic recognition.

In an important sense, a mathematician in any country outside the former Soviet Union is competing with the Soviet émigrés in two distinct markets: The first is a market defined along geographic lines, such as the national job market, where the presence of the émigrés may reduce labor market opportunities (Borjas 2003). The second is a market defined in the realm of ideas, where the mathematician is now competing with Soviets (both the émigrés and the ones left behind) who are doing the same type of research and

who will be competing for journal space and for the rewards that come with increased academic recognition.[55]

Our data can shed light on the relative importance of competition in "bodies" versus "ideas". In particular, the collapse of the Soviet Union led to a moderately sized diaspora of Soviet mathematicians to many countries, not just the United States. To provide a precise definition of the pre-existing mathematics workforce in each country, we restrict the analysis to the years between 1984 and 2008. A mathematician is allocated to a specific country if at least half the papers published between 1984 and 1989 used an affiliation in that country. Table IX lists the 26 countries that had a pre-existing mathematical community of at least 500 active mathematicians prior to the collapse of the Soviet Union. The top 10 countries in this list have at least 2,000 active mathematicians, and account for 72.2 percent of all papers written (outside the Soviet Union) between 1984 and 1989, and these papers received 79.8 percent of all citations. The 26 countries in Table IX account for 89.3 percent of all papers written, and those papers received 93.2 percent of all citations.

We estimate that there are 1,051 mathematicians who published predominantly in the Soviet Union between 1978 and 1989 or who received their doctoral degree from a Soviet institution between 1960 and 1989, yet whose modal affiliation in the post-1992 period was outside the bloc of countries formerly known as the USSR. Of these Soviet mathematicians, 336 (or 32.0 percent) ended up in the United States. Other countries, such as Israel or Germany, also received large numbers of Soviet émigrés (122 in Israel and 66 in Germany). At the same time, however, other countries with a thriving mathematical

---

[55] Kaiser (2012) presents a fascinating analysis of how journals respond to changes in the number of knowledge producers in the market for ideas.

community were barely affected by the migration: China did not receive any Soviet

émigrés, Japan received 7, and India 8.

Table IX also reports the "immigrant share," defined as the ratio (in percent) of the

number of émigrés to the number of pre-existing mathematicians. Not surprisingly, there is

a great deal of variation in the relative size of the supply shock across the 26 countries,

ranging from a maximum of 12.0 percent in Israel, to 5.1 percent in Sweden, 1.9 percent in

Australia, and a minimum of 0.0 percent in China.

There are, of course, quality differences across both the pre-existing mathematicians

in different countries and across the émigrés who chose to settle in different countries. A

conceptually better measure of the supply shock would adjust for these quality differences.

We calculated the percent increase in the number of citations due to the Soviet influx in

each country. We define the quality-adjusted immigrant share as the ratio (in percent) of

the 1984-1989 (AMS) citations received by the émigrés who eventually moved to country i

to the citations received by the mathematicians in that country over the same period. It is

notable that although the quality-adjusted measure of the supply shock differs

substantially from the traditional immigrant share, it is below 3 percent for almost all the

countries. The two outliers are Israel and Sweden, where the Soviet émigrés increased the

effective supply of mathematicians by 10.3 and 16.6 percent, respectively.

We estimated the basic regression model first presented in equation (3) separately

for each of the 26 different countries. The generic regression model is given by:

(7) $\qquad y_{it}(c) = \phi_i + \phi_t + X_i(t)\, \gamma_c + \theta_c\, (T \times Index_i) + \varepsilon_i(t),$

where $y_{ij}(c)$ is the number of papers written by mathematician i, in year t, in country c. We estimated equation (7) using each of the three alternative overlap indices.

The first column of Table IX reports the estimate of the coefficient θ_c (from the specification that uses the index of similarity). There is a lot of dispersion in the estimated coefficient: -1.45 (0.13) in the United States; -0.53 (0.34) in Poland; 0.85 (0.53) in China; and -0.80 (0.66) in Israel. To illustrate the cross-country variation in the measured impact of the collapse of the Soviet Union, Figure IXa shows the scatter diagram relating the estimate of θ_c to the quality-adjusted immigrant share for the 10 countries with the largest mathematical communities.[56] Interestingly, the figure reveals a strong negative correlation between the coefficient θ_c for a particular country and the quality-adjusted immigrant share in that country, at least within the group of the 10 largest countries.

Of course, the impact of the collapse of the Soviet Union on the productivity of mathematicians in any given country depends not only on how many émigrés the country received, but also on how closely "connected" that country is to the global mathematics community. Presumably, mathematicians in a particular country will be more likely to be affected by the post-1992 globalization of the mathematics market if the country's mathematicians are routinely competing in the global market for mathematical ideas.

We construct an index to measure the extent of connectivity between mathematicians in country c and mathematicians in other countries (outside the Soviet Union). Mathematicians in country c published in a specific set of journals in the pre-1990 period. If all of the journals in that set only published articles written by mathematicians

---

[56] The scatter for all 26 countries is less informative since the scaling is skewed by the fact that Israel and, particularly, Sweden have very large quality-adjusted supply shocks.

from country *c*, then country *c* would have a connectivity of 1.0 with itself and 0.0 with all

other countries. But if these journals used half of their slots to publish articles written by

mathematicians from country *c* and the other half to publish articles by mathematicians

from country *d*, then country *c* would have connectivity of 0.5 with itself and 0.5 with

country *d*.

To generalize, let γ_ℓ^c be the share of country *c*'s papers that are published by journal

ℓ (ℓ = 1,...,*L*) and let L_d^ℓ be the share of journal ℓ's papers that are from country *d*. Country

c's index of connectivity with country *d* is then given by the dot product:

(8) $$Y_d^c = \sum_{\ell=1}^{L} \gamma_\ell^c L_d^\ell .$$

The mathematical world is made up by mathematicians living in *C* different countries (κ =

1,...,*C*). We can then define country *c*'s index of connectivity with the mathematical world

as:

(9) $$V_c = \sum_{\kappa=1}^{C} \pi_\kappa Y_\kappa^c$$

where π_κ is the share of the world's population of mathematicians residing in country κ.

We calculated the index defined in equation (9) for each country.[57] The estimated

indices are reported in the last column of Table IX. Note that there is substantial dispersion

in the index of connectivity: It is relatively large for the United States and Canada, but small

for such countries as Belgium, Poland, and Romania. Figure IXb illustrates the relation

between the country-specific coefficient θ_c and the index of connectivity. There seems to be

a negative relation between these variables, at least among the countries with the largest

mathematical communities.

We wish to examine the extent to which the productivity impact of the collapse of

the Soviet Union on competing Western mathematicians (as measured by θ_c) is related to

both the physical shock of émigrés and the increased competition in the global market for

ideas. We estimated a second-stage regression model of the form:

(10) $\theta_c = \alpha_0 + \alpha_1\, p_c + \alpha_2\, V_c + u_c\,,$

where p_c is the quality-adjusted immigrant share, and V_c is the index of connectivity. The

regression is weighted by the inverse of the sampling variance of the dependent variable.

We estimated two alternative specifications of equation (10) to ensure that the

results are not driven by outlying observations. In particular, we estimated the regression

using both the group of the 10 largest countries, as well as the entire group of 26 countries

listed in Table IX. Regardless of the specification used, Table X shows that international

differences in the productivity impact of the collapse of the Soviet Union are strongly and

[57] The data used to calculate the index of connectivity is the full ISI listing of all publications in mathematics
and related fields between 1978 and 1989. Note that these ISI data have not yet been matched with the AMS
database, so that missing papers due to matching difficulties are not a problem.

negatively related to both geographic and idea competition. In other words, the adverse

effect of the collapse of the Soviet Union on competing mathematicians in other countries

was larger in countries that had a large influx of Soviet émigrés and in countries where the

mathematicians traditionally compete in the worldwide market for journal space.

It is also interesting to note that the explanatory power of the second-stage

regressions is quite high; the R^2 is around 0.8 in the regression that uses the 10 countries

with the largest mathematics communities, and between 0.2 to 0.4 in the full set of

countries. It seems that a great deal of the international variation in the productivity impact

of the collapse of the Soviet Union can be explained by a simple model that stresses the

competition introduced by both additional bodies and additional theorems.

VII. Summary

Knowledge production is a central ingredient in any discussion of long-term

economic growth. It is widely recognized that human capital spillovers play a fundamental

role in the knowledge production function: the knowledge produced by one researcher is

both an output and an input into another researcher's production function. Not

surprisingly, there is a consensus among policy makers that increases in the supply of a

highly skilled workforce will increase the productivity of the pre-existing workforce, and

lead to a substantial increase in national wealth.

This paper examines the impact of the influx of renowned Soviet mathematicians

into the American (and global) mathematics community after the collapse of the Soviet

Union. In the period between the establishment and fall of communism, Soviet mathematics

developed in an insular fashion and along different specializations than American

mathematics. After the collapse of the Soviet Union, over 1,000 Soviet mathematicians (nearly a tenth of the pre-existing workforce) migrated to other countries, with about 30 percent settling in the United States. As a result, some fields in the American mathematics community experienced a flood of new mathematicians, theorems, and ideas, while other fields received few Soviet mathematicians and gained few potential insights.

We constructed a data set that contains information on the authorship, subject, and affiliation of all papers published in mathematics over the past 70 years. These data allow us to document the location, affiliation, and complete publication and citation records of mathematicians who were active in either the Soviet Union or the United States for the past few decades.

Our empirical analysis unambiguously documents that the typical American mathematician whose research agenda most overlapped with that of the Soviets suffered a reduction in productivity after the collapse of the Soviet Union. Based solely on the pre-1992 age-output profile of American mathematicians, we find that the actual post-1992 output of mathematicians whose work most overlapped with that of the Soviets and hence could have benefited more from the influx of Soviet ideas is far below what would have been expected. The data also reveal that these American mathematicians became much more likely to switch institutions; that the switch entailed a move to a lower quality institution; that many of these American mathematicians ceased publishing relatively early in their career; and that they became much less likely to publish a "home run" after the arrival of the Soviet émigrés. Although total output declined for the pre-existing group of American mathematicians, the gap was "filled in" by the contribution of Soviet émigrés.

There is no evidence, however, that there was a substantial increase in the size of the American mathematics pie as a result of the Soviet influx.

We also examined how the collapse of the Soviet Union—and the subsequent diaspora of Soviet mathematicians across many countries—differentially affected the productivity of the pre-existing mathematicians in different countries. It turns out that both competition in the local job market (generated by an increase in the number of mathematicians competing for a relatively fixed number of research jobs) and competition in the international journals market (generated by the fact that many mathematicians compete for space in the same limited set of journals regardless of where they reside) help to determine the productivity impact of the collapse of the Soviet Union. The productivity of the mathematicians whose pre-1990 work most overlapped with that of the Soviets declined most when the country received a relatively large number of Soviet émigrés and when that country was strongly interconnected in the journals market with other countries.

Our empirical evidence does not support the conjecture that the Soviet influx generated substantial positive externalities for the pre-existing mathematics workforce. We do not believe this finding arises because American mathematicians did not gain new ideas from the Soviet influx. Rather, we interpret the evidence as suggesting that there may be surprisingly resilient constraints that can counteract these gains. We have shown that some of these constraints exist in the job market and some of the constraints exist in the market for theorems.

There may even be additional constraints that are more fundamental in knowledge production. New research can only be assimilated into the existing body of knowledge if

prominent researchers at the frontier of a field allocate their limited time and effort to do so. As a result, even in a world with nonrival ideas, scarcity of resources and diminishing marginal productivity may come into play.

HARVARD UNIVERSITY AND NATIONAL BUREAU OF ECONOMIC RESEARCH

UNIVERSITY OF NOTRE DAME

Supplementary Data Appendix

Our analysis uses data drawn from three distinct databases: the universe of publications in mathematics as archived by the American Mathematical Society (AMS) Mathematical Reviews, the Thomson Reuters' ISI Web of Science archive, and the Mathematical Genealogy Project (MGP) data that links new doctoral degrees granted in mathematics to their doctoral advisors.

1. AMS Mathematical Reviews

AMS Mathematical Reviews records the titles, publication sources, author names, references, and citations of over 2 million articles from 2,764 different journals and publication sources worldwide. The AMS team (located in Ann Arbor, MI) painstakingly assigns the correct set of articles to each person (even in difficult cases such as identical names), tagging each author with a unique author identifier. The editors of *Mathematical Reviews* also record subject classification codes for each paper in its database: 64 subfields defined according to the 2-digit 2010 Mathematics Subject Classification (MSC2010), 104 subfields (3-digit MSC), or 610 subfields (5-digit MSC). We use the 2-digit coding throughout. Beginning in the early 1980s, the archival record of each published paper provides information not only on the author and the field of the paper, but also on the author's affiliation at the time the paper was published. This institutional information, however, is relatively complete only for papers published since 1984.

The AMS provided us with a spreadsheet containing the publication history of all mathematicians since 1939. A row in this spreadsheet defines a particular permutation of author (i), field (j), and year (t). For each (i, j, t) row we have information on: the number of papers published, the total number of citations received by those papers as of 2011; the institutional affiliation(s) associated with the papers in that row; and the (country) location of the affiliation(s).

Although the count of published papers in these data is complete, the citation information is not. In particular, the AMS database only reports the citations that a paper has received after 2000 (regardless of when the paper was published). In addition, the citation count is based on references made to a paper in a limited list of journals (chosen by the AMS).

2. Thomson Reuters' ISI Web of Science

The ISI Web of Science records the titles, publication source, author names, references, and citations of millions of articles from thousands of journals worldwide. For many articles (especially after 1978), the database records research addresses and reprint addresses for each author, as well as abstracts, keywords, and funding information.

Most important, the database contains complete citation information for each article in a primary set of 7,621 journals, selected to include all of the most important journals in each field. Articles in marginal journals also appear in the database if they either cite an article in the primary database or are cited by such an article. Thus, the ISI Web of Science citation database only misses articles that are not in important journals, are not cited by any articles in important journals, and do not themselves even cite any articles in important journals. This means the citation-adjusted weight of any missed articles must be

very small, especially since ISI adds important journals to its main database by subfields, so even marginal subfields likely have the main field journals represented.

We purchased the records of all 1,179,787 articles in the primary ISI Web of Science database between 1970 and 2009 for the following categories: Mathematics, Applied Mathematics, Interdisciplinary Applications of Mathematics, Mathematical Physics, and Statistics & Probability. We also purchased the records of all 1,921,587 articles referenced by these main articles, and the 2,368,123 papers that cite these main articles. Not counting overlaps in these three categories, this amounted to 3,586,834 articles in Mathematics and related fields. For the reasons mentioned earlier, our ISI database is a relatively complete record of all citations in mathematics and related fields from 1970 until 2009.

We obtained special permission from the AMS to merge our ISI papers by title, source, and author with the AMS database. The AMS set up their own merging site on a mirror in Germany, and we developed software to repeatedly query this mirror with papers from the ISI Web of Science database. We obtained 882,088 matches out of the 1,753,148 journal articles in the AMS database, or just slightly over a 50 percent match rate. The merging allows us to allocate the number of citations to a paper in the ISI database to the matched paper in the AMS database. In addition, the match also allows us to obtain the location and institutional information for additional papers. In particular, the AMS database reports location information on a systematic basis beginning in 1984, while the ISI database provides the same information roughly since 1978.

3. Mathematics Genealogy Project (MGP)

For doctoral degrees awarded in mathematics (since the 14th century), the MGP database reports the name of the degree recipient, the name of the doctoral advisor, the year in which the degree was completed, and the name of the institution granting the degree. In addition to these data, the MGP also provided us with each advisor's and student's AMS unique author identifier. The identifier allows us to merge MGP database with the AMS/ISI databases, and obtain all the available information on the publication and citation records of both students and advisors. The unique AMS identifiers are available for about 65 percent of degrees awarded since 1960.

4. Construction of the sample of predominantly American mathematicians

We focus on a sample of mathematicians whose first published paper occurred on or after 1940, who have fewer than 60 years of "experience" (defined as current year minus the year of first publication), and who were mathematically active prior to the collapse of the Soviet Union (defined as having published at least one paper between 1970 and 1989).

The database provided to us by the AMS reports the mathematician's institutional affiliation and location for any year that the mathematician published at least one paper. We used the AMS data to identify the number of times that the reported institutional affiliation prior to 1989 was located in the United States. The AMS database, however, only began to systematically collect the information on the affiliation associated with any given publication around 1984. We supplement the AMS data by using both the ISI and the MGP data to define the sample of predominantly American mathematicians. As noted above, the ISI data contains affiliation and location information for many other published papers between 1978 and 1983. We include this location information in the time series data for any given mathematician if the equivalent data are not available in the AMS.

The affiliation/location information are sometimes missing in both the AMS and ISI databases (even though the author has published a paper in a particular year). In defining the relative number of times that an author uses an American affiliation, we only include years in which the location information is available.

Finally, the MGP database reports the institution where the mathematician received the doctoral degree. A mathematician is then classified as "predominantly American" if the AMS/ISI location data indicate that more than half of the affiliations of publication prior to 1989 were located in the United States. If the pooled AMS/ISI data do not provide any information on the location of the mathematician prior to 1989, we assume that a mathematician who received a doctoral degree in the United States between 1950 and 1977 is affiliated with an American institution and he, too, is classified as predominantly American.

5. Construction of other samples of mathematicians

The construction of the sample of Soviet mathematicians and Soviet émigrés is analogous to that of predominantly American mathematicians. To the extent possible, we use the AMS data to determine the location of a mathematician prior to 1989, supplement these data with location information from the ISI database for the 1978-1983 period, and further supplement the data with information on the institution from which the mathematician received the doctoral degree. The universe of active Soviet mathematicians as of 1989 is defined as the sample of persons who: (a) published at least one paper between 1970 and 1989; and (b) published at least half of the pre-1989 papers using a Soviet affiliation, or, in the absence of any location information from the AMS/ISI data, received his or her doctoral degree from a Soviet institution after 1950.

To determine the "émigré" status of a Soviet mathematician, we must observe the mathematician publishing at least one paper between 1992 and 2008. A Soviet mathematician is an émigré if the modal affiliation of publication between 1992 and 2008 is located outside the geographic boundaries of the former Soviet Union and is a valid country code in the sense that it provides an actual location for an institution. The country of destination for the Soviet émigré is defined in terms of the modal affiliation used by the mathematician between 1992 and 2008.

Finally, the study of international differences defines a sample of pre-existing mathematicians for each of 26 countries. We use the affiliation data available in the AMS beginning in 1984 to classify the mathematicians into the various countries. An active pre-existing mathematician in country c is someone who published at least one paper between 1984 and 1989, and whose modal affiliation in that period was located in that country.

6. Construction of the number of "home runs"

To estimate the number of home runs published by each American mathematician in each year, we first calculate the total number of worldwide lifetime citations received by each article in the ISI database that has been successfully matched to the AMS database. Note that these lifetime citations are complete for those articles that were published in the base set of journals fully covered by ISI; for articles that cite these journals and are cited by these journals, the lifetime citations only include citations made by articles in the base set of journals. We then used these data to calculate the 90th percentile, 95th percentile, and 99th percentile thresholds of the lifetime citations distribution for papers published in each

year. We use the matched AMS/ISI database to count the number of papers that a particular mathematician published each year that exceed these threshold percentiles.

References

Abramitzky, Ran, and Isabelle Sin. "Book Translations as Idea Flows: The Effects of the Collapse of Communism on the Diffusion of Knowledge," Stanford University Working Paper, 2011.

American Mathematical Society, "Academic Recruitment and Hiring Report," *AMS Notices*, various issues.

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang. "Superstar Extinction," *Quarterly Journal of Economics*, 125 (2010), 549-589.

Borjas, George J. "The Labor Demand Curve *Is* Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *Quarterly Journal of Economics*, 118 (2003), 1335-1374.

Foray, Dominique. *Economics of Knowledge* (Cambridge: MIT Press, 2004).

Furman, Jeff, Margaret Kyle, Iain Cockburn, and Rebecca Henderson. "Knowledge Spillovers, Geographic Location and the Productivity of Pharmaceutical Research," *Annales d'Economie et de Statistique*, 79/80 (2005), 167-190.

Ganley, Gladys D. "Unglued Empire: The Soviet Experience with Communications Technologies," (New York: Praeger, 1996).

Ganguli, Ina. "Saving Soviet Science: The Impact of Grants When Government R&D Funding Disappears," unpublished manuscript, Harvard University, 2010.

Graham, Loren, and Jean-Michel Kantor. *Naming Infinity: A True Story of Religious Mysticism and Mathematical Creativity* (Cambridge: Harvard University Press, 2009).

Howe, Marvine. "For Émigré Scientists, Job Hunting is Difficult," *The New York Times*, August 12, 1990.

Hunt, Jennifer, and Marjolaine Gauthier-Loiselle. "How Much Does Immigration Boost

Innovation?" *American Economic Journal: Macroeconomics*, 2 (2010), 31–56.

International Monetary Fund et al, *A Study of the Soviet Economy*, Volume 3 (Washington, DC:

IMF, 1991).

Jones, Charles I. "Growth and Ideas," *Handbook of Economic Growth*, 1 (2005), 1063-1111.

Jones, Charles I., and Paul M. Romer. "The New Kaldor Facts: Ideas, Institutions, Population,

and Human Capital," *American Economic Journal: Macroeconomics*, 2 (2010), January

2010.

Kaiser, David. "Booms, Busts, and the World of Ideas: Enrollment Pressures and the Challenge

of Specialization," *Osiris*, forthcoming 2012.

Kolata, Gina. "Soviet Scientists Flock to U.S., Acting as Tonic for Colleges," *The New York Times*,

May 8, 1990.

Laqueur, Walter. *The Dream That Failed: Reflections on the Soviet Union* (New York: Oxford

University Press, 1996).

Lefschetz, Solomon, "Topology," in *Summary Report of the Panel on Soviet Mathematics,* Joseph

P. LaSalle, ed. (Baltimore: Research Institute for Advanced Study, 1961).

Lucas, Robert E. "On the Mechanics of Economic Development," *Journal of Monetary

Economics*, 22 (1988), 3-42.

Lucas, Robert E. "Ideas and Growth," *Economica*, 76 (2009), 1-19.

McClure, Donald. "AMS Task Force on Employment Report to the Mathematical Community

Part I: Academic Hiring Survey, 1991-1992." *AMS Notices*, 39 (1992), 754-758.

O'Connor, J.J., and E.F. Robertson, "Biography of Nikolai Nikolaevich Luzin," Mac Tutor History of Mathematics Archive, University of St. Andrews, Scotland, 1999. http://www-history.mcs.st-and.ac.uk/.

O'Dette, Ralph E. "Russian Translation," *Science*, 125 (1957), 579-585.

Polyak, B.T. "History of Mathematical Programming in the USSR: Analyzing the Phenomenon," *Mathematical Programming*, 91 (2002), 401-416.

Romer, Paul M. "Increasing returns and long-run growth," *Journal of Political Economy*, 1986.

Romer, Paul M. "Endogenous Technological Change," *Journal of Political Economy*, Vol. 98, No. 5, Part 2 (Oct., 1990), pp. S71-S102.

Smithies, F. "The 1936 Luzin Affair: Review of *Delo akademika Nikolaya Nikolaevicha Luzina* [The Case of Academician Nikolai Nikolaevich Luzin]," Mac Tutor History of Mathematics Archive, University of St. Andrews, Scotland, 2003. http://www-history.mcs.st-and.ac.uk/.

Tybulewicz, A. "Cover-to-Cover Translations of Soviet Technical Journals." *Aslib Proceedings*, 22 (1970), 55-62.

U.S. National Science Foundation. *Federal Funds for Research and Development: Fiscal Years 1973–2003*. (Washington: Division of Science Resources Statistics, 2004).

Waldinger, Fabian. "Quality Matters: The Expulsion of Professors and the Consequences for Ph.D. Student Outcomes in Nazi Germany," *Journal of Political Economy*, 118 (2010), 787-831.

Waldinger, Fabian. "Peer Effects in Science—Evience from the Dismissal of Scientists in Nazi Germany," *Review of Economic Studies*, forthcoming 2012.

Figure I

Ratio of Soviet papers to American papers, by field, 1984-89
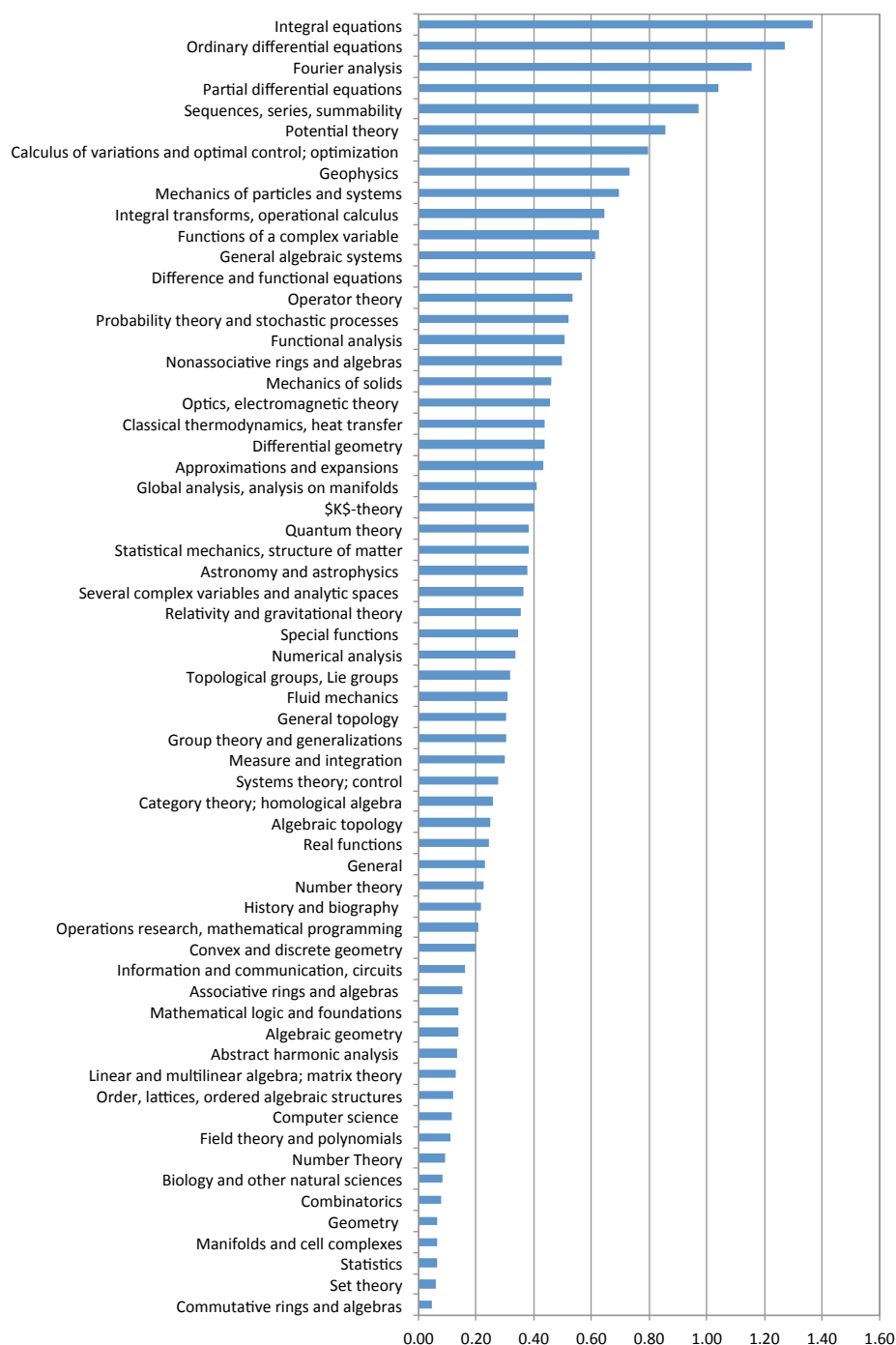
Notes: The paper counts are obtained directly from the web-based AMS Mathematical Reviews database (*MathSciNet*) and consist of all papers published by mathematicians affiliated with Soviet or American institutions between 1984 and 1989.

Figure II

Trend in coauthorship rate between Soviet and American mathematicians
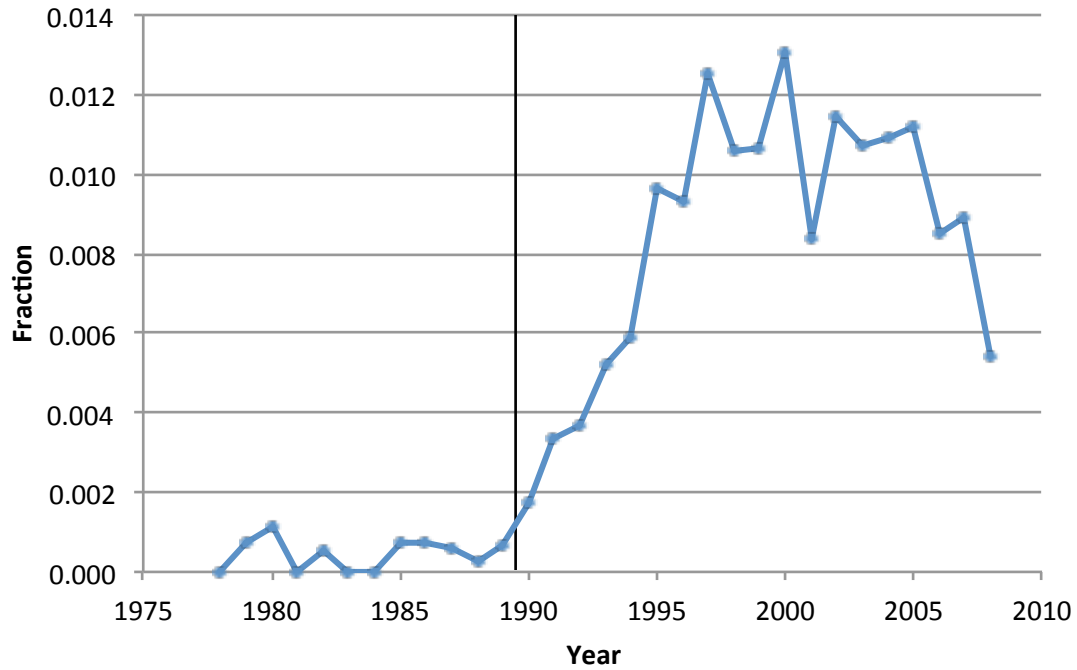
Notes: The denominator of this fraction is the number of papers published each year where at least one author reports an American affiliation. The numerator is the number of such papers in which one other author also reports a Soviet affiliation.

Figure III

**Employment trends for new mathematics doctorates
granted by North American institutions**
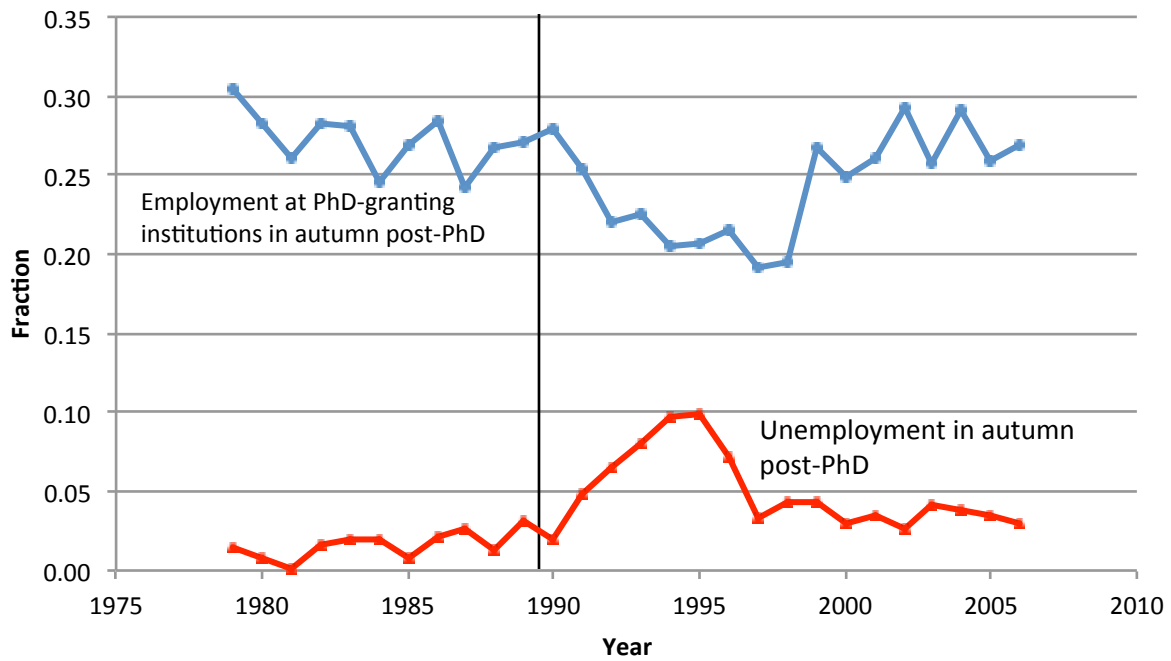
Source: Data compiled by the authors from American Mathematical Society (various issues).

Figure IV. Citations to American or Soviet papers by American mathematicians

A. Citations to American papers
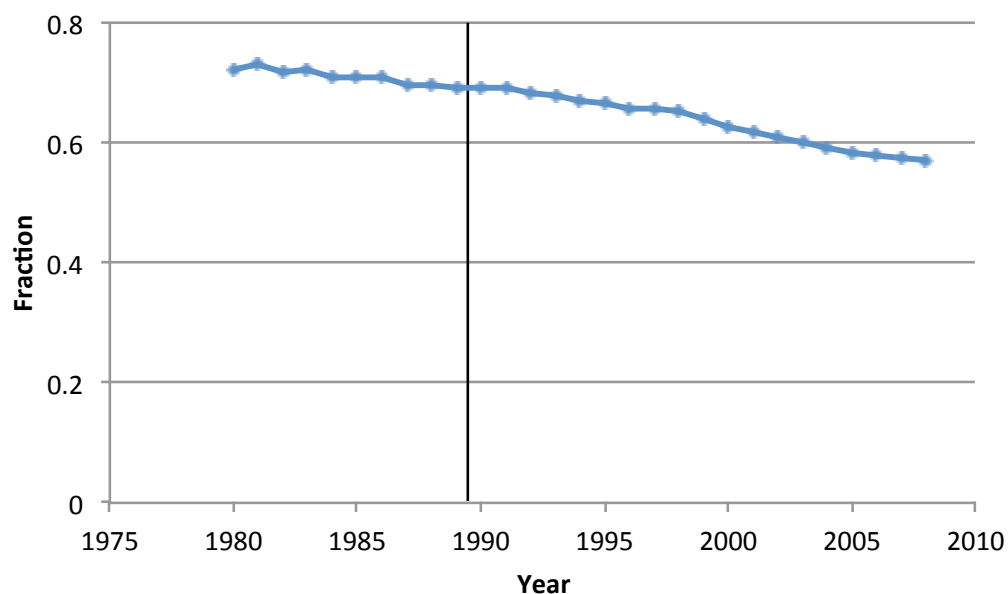
B. Citations to Soviet papers
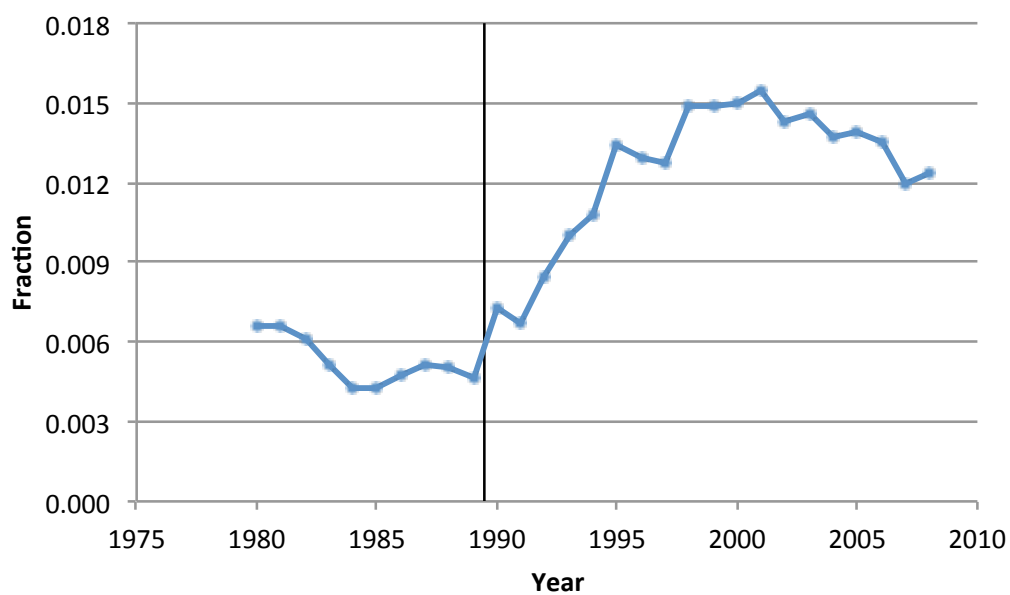
Notes: The numerator is the number of citations in American papers published in a given year that go to either American (panel A) or Soviet (panel B) papers published during the entire 1978-2008 period. The denominator is the total number of citations made by American papers in any given year to papers published with a valid research address between 1978 and 2008.

Figure V

Frequency distribution of output for American mathematicians

a. Papers
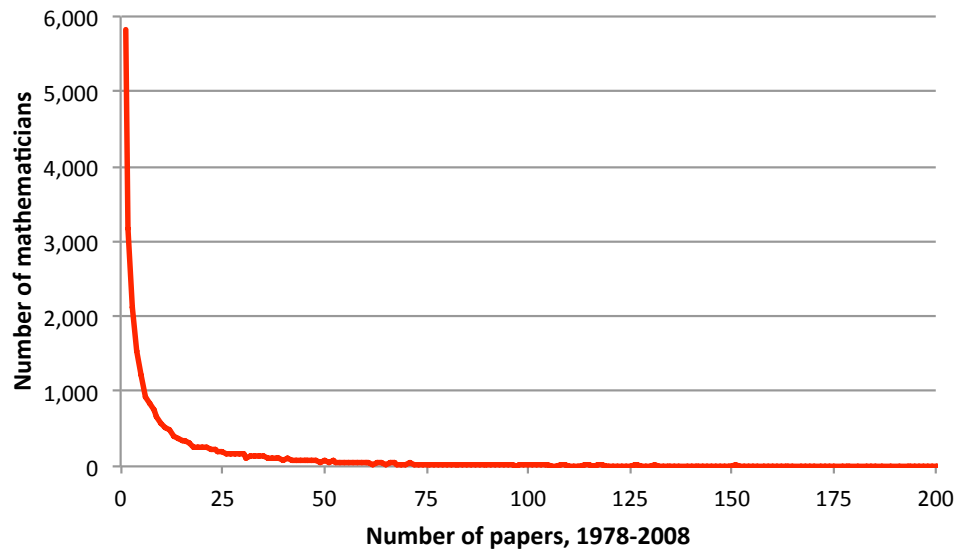
b. Citations (ISI)
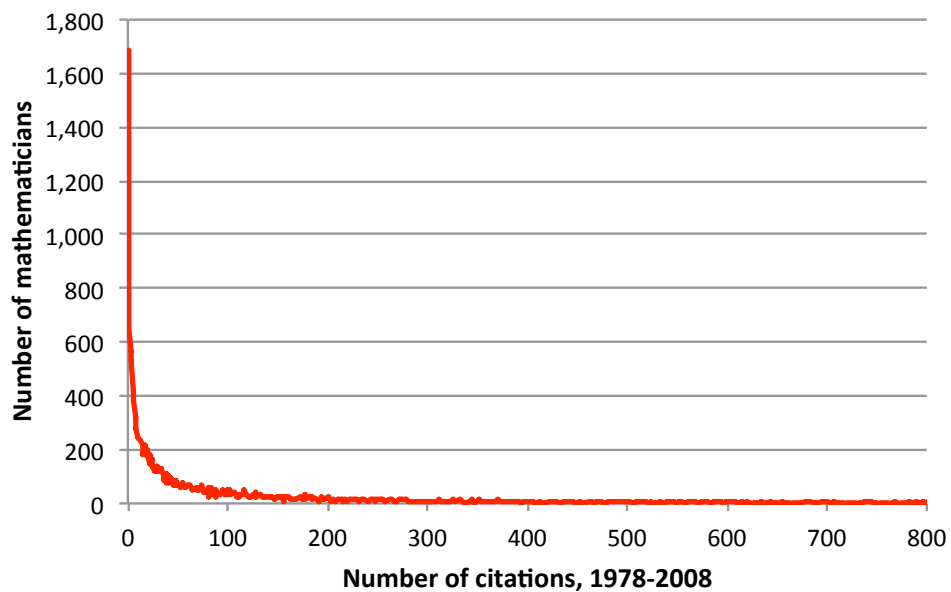
Note: The sample consists of mathematicians who were predominantly affiliated with an American institution before 1990 and who published at least one paper between 1978 and 2008. The maximum number of papers is 801, and the maximum number of (ISI) citations is 31,962.

Figure VI

Fraction of publications published by Soviet émigrés, by type of field
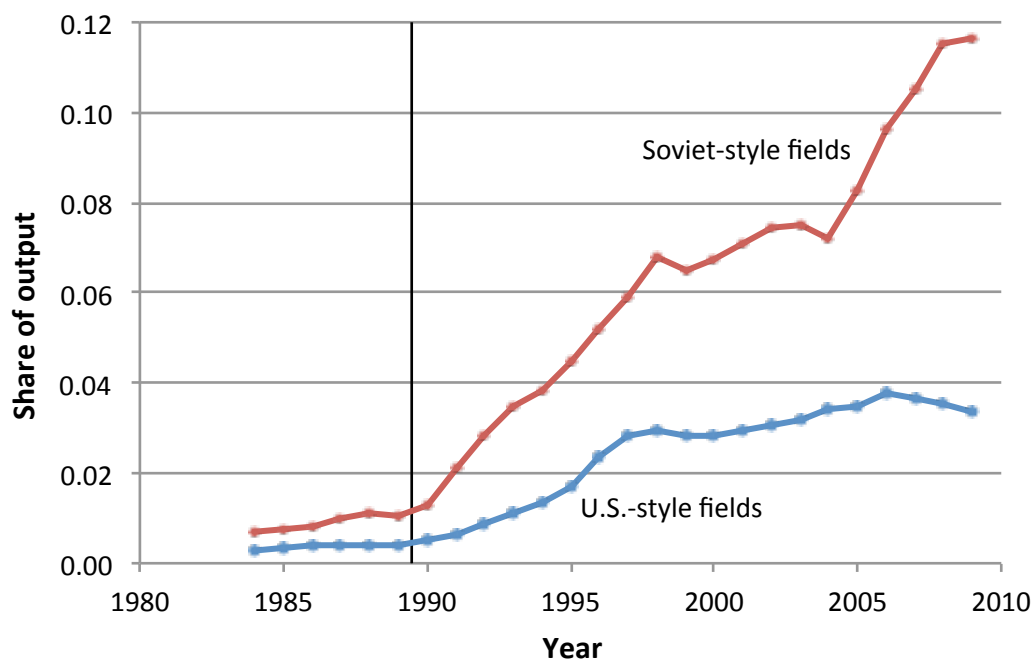
Notes: The U.S.-style fields consist of the bottom 10 fields in Figure I (those with the highest ratios of pre-influx American papers to pre-influx Soviet papers), while the Soviet-style fields consist of the top 10 fields in Figure 1 (those with the lowest ratios). The denominator of the "share of output" is the number of papers in the given subfields published in the United States by mathematicians who were never affiliated with a Soviet institution. The numerator is the number of papers in the given subfield published in the United States by mathematicians who were active before 1990 and had a former Soviet affiliation.

Figure VII

Impact of index of similarity on output of American mathematicians

A. Annual number of papers per mathematician
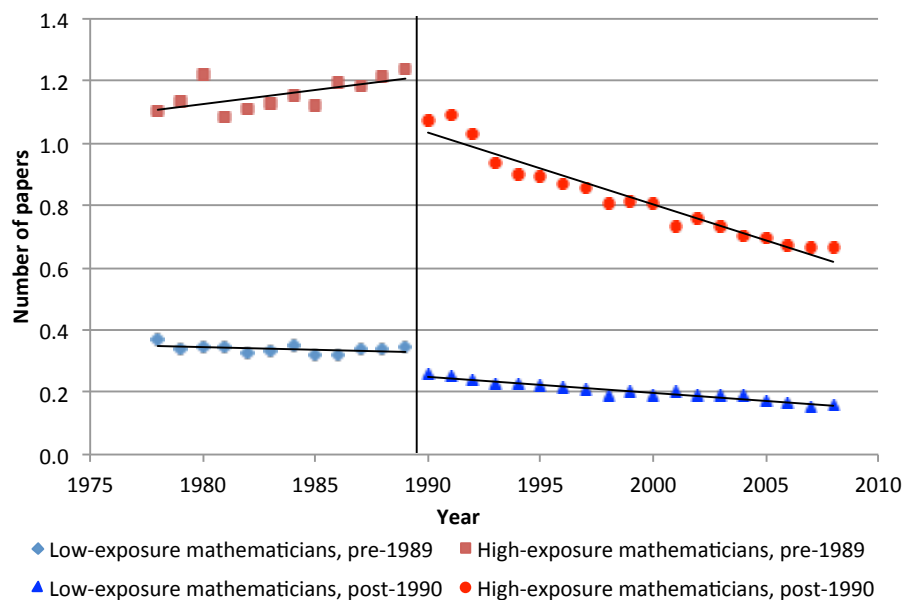
b. Annual number of papers per mathematician, removing individual fixed effects
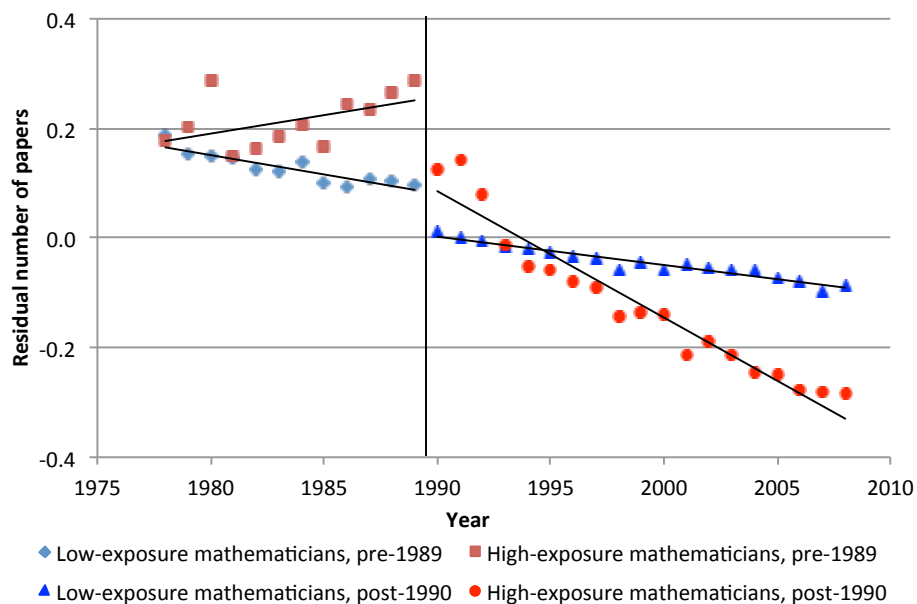
Notes: The low exposure group consists of mathematicians in the bottom quartile of the distribution of the index of similarity, while the high exposure group consists of mathematicians in the top quartile. The residual papers in panel B are calculated from a regression that contains individual fixed effects (demeaning the data for each individual).

Figure VIII

Survival functions for American mathematicians, by level of exposure

a. Low-exposure mathematicians
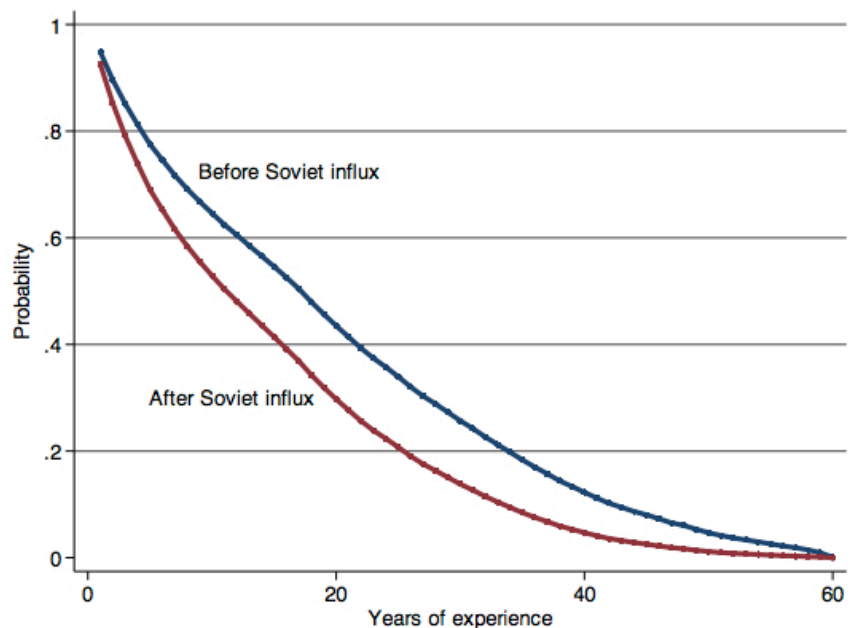
b. High-exposure mathematicians
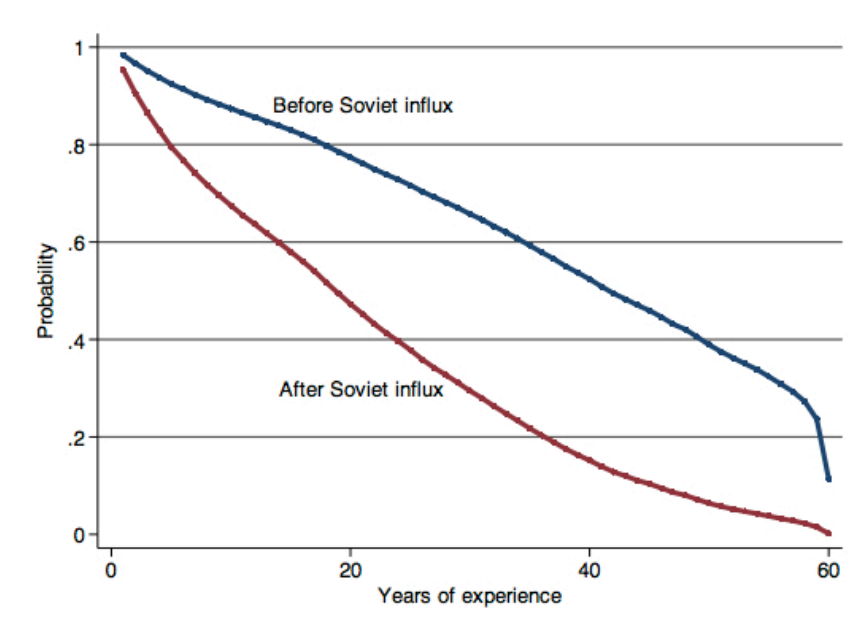
Notes: The survival curve is calculated from the hazard function specification reported in row 1, column 3 of Table 4. A low (high) exposure mathematician is the typical mathematician in the bottom (top) quartile of the distribution of the index of similarity.

Figure IX

**International differences in productivity impact of Soviet influx
(countries with more than 2,000 mathematicians)**

a. Immigrant share
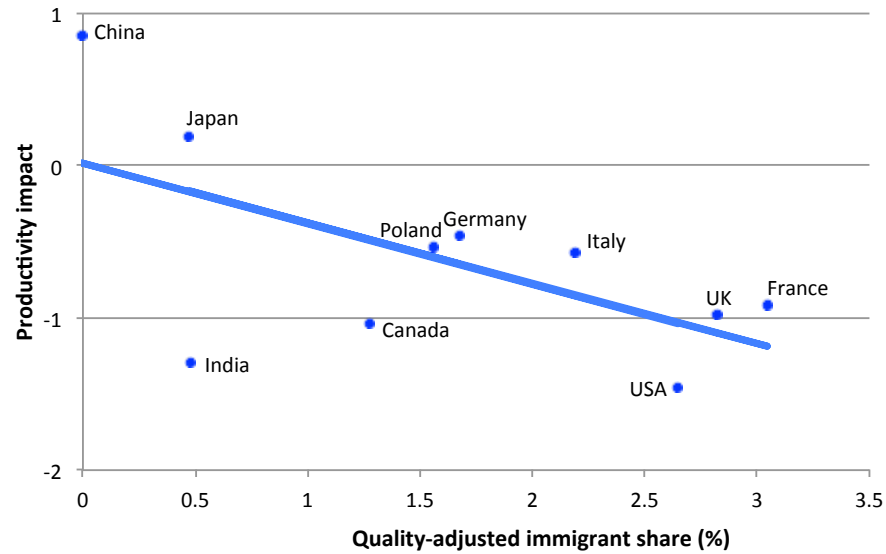
b. Journal connectivity
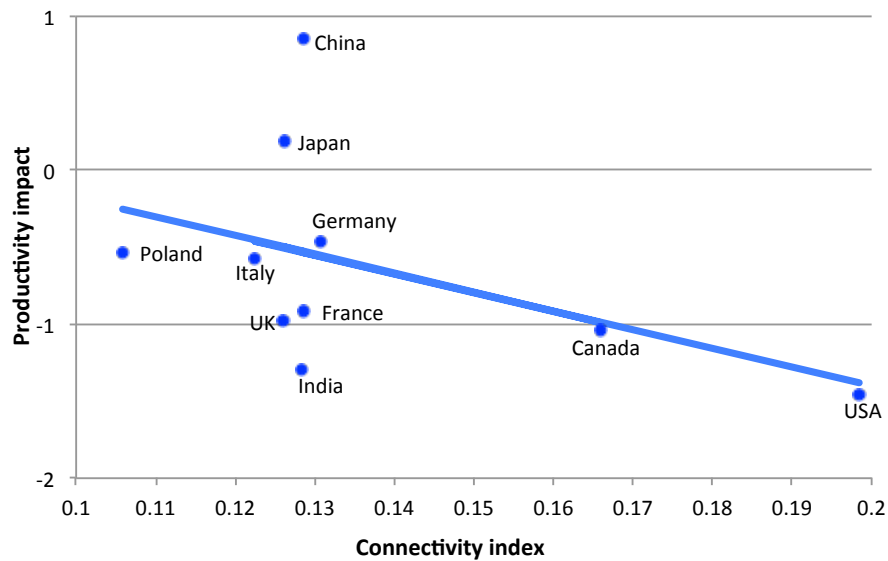
Notes: The productivity impact is the coefficient of the interaction between the index of similarity and the post-1992 dummy variable from a first-stage regression estimated separately in each country. The dependent variable in the first-stage regression is the number of papers published by a mathematician in a given year. The quality-adjusted immigrant share is the ratio of the 1984-1989 citations received by the émigrés who eventually moved to a particular country to the number of 1984-1989 citations generated by the pre-existing mathematicians in that country. The index of connectivity is defined in the text.

Table I

Summary statistics for samples of American and Soviet mathematicians

Variable:	Americans	Soviet émigrés to US	Soviet émigrés elsewhere	All other Soviets
Number of mathematicians	29392	336	715	11173
Papers published, 1978-1991				
Mean papers per mathematician	6.7	17.8	14.6	8.1
Median papers	3.0	13.0	10.0	5.0
Maximum number of papers	232.0	104.0	152.0	180.0
Papers published, 1992-2008				
Mean papers per mathematician	6.8	27.2	28.8	7.6
Median papers	1.0	21.0	22.0	1.0
Maximum number of papers	768.0	128.0	317.0	311.0
Citations, AMS, 1978-1991				
Mean citations per mathematician	29.1	74.6	32.8	8.6
Median citations	1.0	10.0	6.0	0.0
Maximum number of citations	5550.0	1276.0	1441.0	2928.0
Citations, AMS, 1992-2008				
Mean citations per mathematician	33.6	177.4	110.3	13.4
Median citations	0.0	62.0	37.0	0.0
Maximum number of citations	3404.0	1709.0	1988.0	1287.0
Citations, ISI, 1978-1991				
Mean citations per mathematician	110.2	185.1	79.8	25.3
Median citations	20.0	25.5	11.0	3.0
Maximum number of citations	20274.0	7232.0	3040.0	3054.0
Citations, ISI, 1992-2008				
Mean citations per mathematician	52.1	209.0	156.2	27.3
Median citations	0.0	88.5	60.0	0.0
Maximum number of citations	11688.0	3371.0	4442.0	1258.0
Median number of fields	2.0	5.5	5.0	2.0
Percent first published after 1980	45.2	40.5	46.7	48.8

Notes: The sample consists of "active" mathematicians who published at least one paper between 1970 and 1989. An American mathematician is someone who was predominantly affiliated with an American institution prior to 1989, while a Soviet mathematician is someone who was predominantly affiliated with a Soviet institution prior to 1989. The post-1992 modal affiliation of a Soviet émigré is located outside the geographic boundaries of the former USSR.

Table II

American mathematicians with the largest number of publications, 1960-89

Mathematician:	Number of papers	Correlation coefficient	Index of intensity	Index of similarity
Frank Harary	416	-0.039	0.101	0.096
Pranab Kumar Sen	318	0.035	0.199	0.090
Richard Ernest Bellman	317	0.420	0.339	0.410
Ciprian Foias	299	0.295	0.414	0.291
Avner Friedman	290	0.629	0.633	0.411
Robert E. Kalaba	241	0.272	0.306	0.308
Peter C. Fishburn	232	0.157	0.302	0.126
Bang-Yen Chen	220	0.009	0.164	0.070
Barry Simon	217	0.537	0.580	0.325
V. Lakshmikantham	215	0.467	0.633	0.310

Notes: The correlation coefficient, index of intensity, and index of similarity measure each mathematician's research overlap with the Soviet research program for papers published between 1960 and 1989. See the text for the definition of the various indices.

Table III

Impact of Soviet supply shock on American mathematicians

Specification/regressor	Mathematicians predominantly in U.S.		Mathematicians always in U.S.	
	Number of papers	Number of citations	Number of papers	Number of Citations
A. Author-year regressions				
Correlation coefficient	-0.133	-19.577	-0.116	-16.298
	(0.036)	(1.576)	(0.034)	(1.540)
Index of intensity	-0.047	-14.845	-0.042	-12.293
	(0.028)	(1.293)	(0.027)	(1.261)
Index of similarity	-1.523	-69.155	-1.419	-58.494
	(0.113)	(4.645)	(0.108)	(4.655)
B. Author-year regressions, short run				
Correlation coefficient	-0.102	-14.214	-0.085	-11.404
	(0.032)	(1.783)	(0.030)	(1.410)
Index of intensity	-0.045	-10.944	-0.039	-8.830
	(0.023)	(1.221)	(0.022)	(1.181)
Index of similarity	-1.056	-48.547	-0.985	-39.054
	(0.111)	(4.232)	(0.108)	(4.117)
C. Author-year regressions, long run				
Correlation coefficient	-0.122	-25.219	-0.108	-21.095
	(0.049)	(2.037)	(0.046)	(2.019)
Index of intensity	-0.019	-19.179	-0.015	-15.889
	(0.039)	(1.687)	(0.037)	(1.666)
Index of similarity	-1.930	-91.211	-1.802	-77.930
	(0.150)	(5.961)	(0.145)	(6.055)
D. Author-field-year regressions				
Correlation coefficient	-0.0021	-0.3048	-0.0020	-0.2578
	(0.0006)	(0.0249)	(0.0005)	(0.0244)
Index of intensity	-0.0007	-2378	-0.0007	-0.2005
	(0.0004)	(0.0206)	(0.0004)	(0.0202)
Index of similarity	-0.0238	-1.0248	-0.0240	-0.8696
	(0.0017)	(0.0732)	(0.0016)	(0.0732)

Notes: Standard errors are reported in parentheses and clustered at the individual level. The table reports the coefficient of the interaction between the overlap index and a post-1992 dummy variable. Each observation in panels A, B, and C represents a unique author-year permutation; an observation in panel D represents a unique author-field-year permutation. The regressions use the entire sample period 1978-2008 in Panels A and D, the 1978-1999 period in the short-run, and the 1978-1991 and 2000-2008 periods in the long run. The numbers of papers (citations) regressions in the predominantly American sample have 804,180 (611,916) observations in Panel A; 540,896 (413,082) observations in Panel B; 510,260 (389,836) observations in Panel C; and 51,467,520 (37,278,675) observations in Panel D. All regressions include the mathematician's years of experience (introduced as a quartic polynomial), year fixed effects, and individual fixed effects. The regressions in panel D also include all interactions between the year fixed effects and the field fixed effects.

Table IV

Impact of Soviet supply shock on number of "Home Runs"
(Conditional fixed effects Poisson model)

| Sample: | Number of papers with citations above x^{th} percentile: | | | By coauthorship status (95th percentile) | |
	90th	95th	99th	Coauthored with émigré	Did not coauthor
Correlation coefficient in:					
Bottom quartile	0.058	-0.055	0.211	0.005	-0.055
	(0.066)	(0.097)	(0.237)	(0.279)	(0.103)
Middle 50 percent	0.012	0.029	0.120	0.058	0.021
	(0.032)	(0.044)	(0.090)	(0.144)	(0.046)
Top quartile	-0.204	-0.160	-0.121	0.014	-0.215
	(0.032)	(0.043)	(0.087)	(0.091)	(0.049)
Index of intensity in:					
Bottom quartile	0.089	-0.025	0.127	0.022	-0.025
	(0.068)	(0.101)	(0.260)	(0.286)	(0.108)
Middle 50 percent	-0.043	-0.025	0.112	0.040	-0.036
	(0.031)	(0.042)	(0.086)	(0.131)	(0.044)
Top quartile	-0.166	-0.123	-0.115	0.019	-0.169
	(0.034)	(0.045)	(0.090)	(0.095)	(0.051)
Index of similarity in:					
Bottom quartile	0.331	0.366	0.402	-0.190	0.393
	(0.069)	(0.095)	(0.198)	(0.507)	(0.096)
Middle 50 percent	0.011	-0.020	-0.004	-0.096	-0.024
	(0.037)	(0.050)	(0.106)	(0.162)	(0.053)
Top quartile	-0.158	-0.122	0.016	0.072	-0.180
	(0.029)	(0.039)	(0.079)	(0.085)	(0.044)

Notes: Standard errors reported in parentheses. The sample consists of predominantly American mathematicians. An observation consists of an author-year pair for each year between 1978 and 2005. The dependent variable is the number of "home runs", defined as the number of papers written in that year that have received a sufficiently high number of citations to place them in the top 90th, 95th, or 99th percentile for papers published that year. The regression includes a post-1992 dummy variable, years of experience (introduced as a quartic), and individual fixed effects. The coefficient reported in the table is that of the post-1992 dummy variable. Each coefficient in the table is drawn from a separate Poisson regression estimated in the respective sample. The regressions in the last two columns re-estimate the 95th percentile regressions separately on the samples of American mathematicians who either coauthored or did not coauthor with a Soviet émigré at any point after 1992. The number of observations for the (bottom, middle, top) groups are (142,563, 292,446, 156,996) in Panel A, (136,590, 303,751, 151,664) in Panel B, and (128,617, 289,149, 174,239) in Panel C.

Table V

**Impact of Soviet supply shock on probability of "retirement" from publishing
(Cox proportional hazard models)**

Sample	Measure of overlap		
	Correlation coefficient	Index of intensity	Index of similarity
All pre-existing mathematicians	0.410 (0.090)	0.230 (0.084)	5.571 (0.298)
Less than 10 years of experience	1.099 (0.229)	0.653 (0.176)	10.340 (0.962)
10-19 years of experience	0.166 (0.192)	0.299 (0.175)	0.232 (0.645)
At least 20 years of experience	0.099 (0.181)	0.101 (0.183)	1.433 (0.491)

Notes: Standard errors are reported in parentheses and are clustered at the individual level. The sample consists of predominantly American mathematicians. Each observation in the data represents an individual-year of experience record, beginning in 1978 until the year of the individual's last observed publication. The regression model includes a post-1992 dummy variable, the mathematician's overlap index, and the interaction between the overlap index and the post-1992 indicator (which is the coefficient reported in the table). The sample sizes (by row) are: 446,720 observations, 151,012 observations, 143,442 observations, and 152,267 observations.

Table VI

Impact of Soviet supply shock on institutional mobility

	Measure of overlap/Dependent variable					
	Correlation coefficient		Index of intensity		Index of similarity	
Sample/regressor	Moved	Δ Quality	Moved	Δ Quality	Moved	Δ Quality
A. All mathematicians						
Institution hired émigré	0.046	-2.382	0.046	-2.383	0.047	-2.385
	(0.013)	(.122)	(0.013)	(.122)	(0.013)	(.122)
Overlap index	0.172	-0.415	0.158	-0.282	0.321	-1.329
	(0.025)	(0.308)	(0.022)	(0.252)	(0.066)	(.997)
B. First published after 1980						
Institution hired émigré	0.098	-2.408	0.098	-2.411	0.099	-2.407
	(0.020)	(.158)	(0.020)	(.158)	(0.020)	(.158)
Overlap index	0.216	-0.618	0.157	-0.394	0.676	-2.606
	(0.040)	(0.373)	(0.031)	(0.285)	(0.151)	(1.499)
C. First published before 1980						
Institution hired émigré	0.001	-2.274	0.001	-2.277	0.001	-2.272
	(0.015)	(.188)	(0.015)	(.189)	(0.015)	(.189)
Overlap index	0.100	0.225	0.099	0.356	0.218	-.180
	(0.029)	(0.538)	(0.029)	(0.540)	(0.069)	(1.345)

Notes: Robust standard errors are reported in parentheses. The sample consists of predominantly American mathematicians whose last observed affiliation before 1990 and the modal post-1992 observation were both in the United States. The dependent variable in the “moved” regressions is a dummy variable set to unity if the mathematician’s last observed affiliation before 1990 is different from the post-1992 modal affiliation, while the dependent variable in the “change in quality” regressions is the difference (among movers) in the log number of papers published by these two institutions. The “moved” regressions have 13,137 observations in panel A, 6,513 observations in panel B, and 6,624 observations in panel C. The “change in quality” regressions have 4,029 observations in panel A, 3,021 observations in panel B, and 1,008 observations in panel C. All regressions hold constant the log number of papers published by the mathematician between 1978 and 1990 and the mathematician’s years of work experience; the “moved” regressions also include the log number of papers published by the pre-1990 institution.

Table VII

Impact of Soviet supply shock on productivity of students
(Dependent variable = Student's annual number of papers or citations)

Index/Variable	Papers			Citations (ISI)		
	(1)	(2)	(3)	(1)	(2)	(3)
Correlation coefficient						
Soviet advisor	0.225	0.232	---	0.557	0.693	---
	(0.040)	(0.041)		(0.211)	(0.223)	
Advisor's index × T	-0.425	-0.386	-0.178	-6.936	-6.482	-3.808
	(0.092)	(0.091)	(0.106)	(1.124)	(1.127)	(1.445)
Index of intensity						
Soviet advisor	0.256	0.265	---	1.147	1.309	---
	(0.041)	(0.042)		(.225)	(.236)	
Advisor's index × T	-0.508	-0.465	-0.236	-9.036	-8.495	-5.429
	(0.095)	(0.095)	(0.106)	(1.346)	(1.349)	(1.576)
Index of similarity						
Soviet advisor	0.295	0.296	---	0.875	0.877	---
	(0.041)	(0.042)		(0.223)	(0.235)	
Advisor's index × T	-0.408	-0.369	-0.129	-5.529	-5.143	-1.302
	(0.187)	(0.182)	(0.242)	(1.944)	(1.976)	(3.073)
Institution fixed effects	No	Yes	Yes	No	Yes	Yes
Advisor's fixed effects	No	No	Yes	No	No	Yes

Notes: Robust standard errors are reported in parentheses. The dependent variable is the average annual number of papers published or (ISI) citations received between the time the student obtained his doctoral degree and 2011. The sample consists of students who received doctoral degrees from American institutions between 1978 and 2008, and whose advisor was either a predominantly American pre-existing mathematician or a Soviet émigré. The papers regressions have 20,822 observations, and the citations regressions have 20,743 observations. The "Soviet Advisor" variable is a dummy variable set to unity if the advisor is a Soviet émigré; the "Advisor index" variable is the American's advisor's value of the overlap index (set to zero if the advisor is a Soviet émigré); and T is a dummy variable set to unity for the post-1992 period. All regressions include the value of the advisor's overlap index (if American) and a set of fixed effects indicating the calendar year of receiving the doctoral degree.

Table VIII

Predicting annual output of American mathematicians

| | American mathematicians | | | | Soviet | |
	Most exposed	Middle 50 percent	Least exposed	All Americans	émigrés	Total
Number of papers, short-run prediction						
Actual	4015.0	3606.6	1253.3	8874.9	370.6	9245.5
Predicted	5062.0	3519.2	900.6	9481.8	---	9481.8
Difference	-1047.0	87.4	352.7	-606.9	---	-235.9
	(121.2)	(126.5)	(72.1)	(189.5)		(189.5)
Long-run prediction: Number of papers						
Actual	3350.6	3013.1	1035.3	7398.9	432.5	7831.4
Predicted	3879.9	2910.3	757.4	7547.5	---	7547.5
Difference	-529.3	102.8	277.9	-148.6	---	283.9
	(76.3)	(119.0)	(88.0)	(166.5)		(166.5)
Number of citations						
Actual	19355.5	13372.8	3939.2	36667.5	2944.8	39612.3
Predicted	22088.5	13191.7	2914.3	38194.5	---	38194.5
Difference	-2733.0	181.1	1024.9	-1527.0	---	1417.8
	(1083.2)	(1693.2)	(1239.9)	(2361.7)		(2361.7)

Notes: The standard errors reported in parentheses give the forecast error of the predicted number of papers or citations, adjusted for the correlation across observations for the same individual. The sample consists of persons who were always affiliated with an American institution prior to 1990. The predicted number of papers in the short run is based on a regression estimated separately in each of the three samples of mathematicians ranked according to their placement in the distribution of the coefficient of similarity. The unit of observation is an author-year; the regression uses observations between 1970 and 1991; and the regressors include a quartic in experience and individual fixed effects. We then use this regression to predict each mathematician's output in each year between 1992 and 1999, and the table reports the sum of these predictions. The long-run predictions are based on the specification estimated over the entire sample period 1978-2008 and reported in Panel A of Table 3 (using the index of similarity). This regression is then used to predict a mathematician's output in each year between 1992 and 2008 assuming that no Soviet influx occurred (so the index of similarity plays no differential role before and after 1992). The predictions are then summed within each of the three exposure groups. The counts of papers and citations of Soviet émigrés only include the post-1992 papers published and citations received while the émigré was affiliated an American institution.

Table IX

The diaspora of Soviet mathematicians after the collapse of the Soviet Union

Country:	First-stage coefficient	Number of mathematicians	Number of émigrés	Immigrant share (%)	Quality-adjusted share (%)	Connectivity index
United States	-1.452**	22449	336	1.50	2.65	0.199
China	0.853*	6563	0	0.00	0.00	0.129
Germany	-0.462	6223	66	1.06	1.68	0.131
Japan	0.190	4848	7	0.14	0.47	0.126
France	-0.915**	4114	72	1.75	3.05	0.129
United Kingdom	-0.982**	3804	78	2.05	2.82	0.126
Italy	-0.577**	3364	12	0.36	2.19	0.123
Canada	-1.040**	3155	36	1.14	1.27	0.166
India	-1.299**	3098	8	0.26	0.48	0.129
Poland	-0.530*	2310	33	1.43	1.56	0.106
Spain	1.902**	1614	7	0.43	0.20	0.133
Netherlands	-1.748**	1374	7	0.51	0.06	0.143
Australia	0.459	1178	22	1.87	1.64	0.139
Israel	-0.803	1014	122	12.03	10.33	0.169
Romania	-0.177	1012	3	0.30	0.39	0.103
Czechoslovakia	-0.435	933	4	0.43	3.09	0.094
Brazil	0.132	822	16	1.95	1.52	0.145
Bulgaria	-0.108	808	4	0.50	0.16	0.133
Switzerland	-0.942*	661	4	0.61	0.10	0.130
Belgium	-1.401*	655	19	2.90	1.22	0.091
Hungary	-2.954**	654	5	0.76	0.53	0.117
Taiwan	-2.043	590	1	0.17	1.71	0.155
Yugoslavia	-0.612	543	2	0.37	0.00	0.119
Greece	-1.540	539	1	0.19	0.25	0.141
Austria	-0.387	520	5	0.96	0.23	0.128
Sweden	-1.497**	513	26	5.07	16.61	0.135

Notes: An "active" mathematician published at least one paper in the period between 1984 and 1989. The active mathematicians are then allocated to the country of their modal affiliation during this period. The first-stage coefficient is the coefficient of the interaction between the index of similarity and the post-1992 dummy variable, estimated separately in each country; an "*" indicates that the coefficient is between 1.50 and 1.99 times the size of the clustered standard error, and "**" indicates that it is at least twice the size of the standard error. The destination country for the Soviet émigrés is the modal country of affiliation between 1992 and 2008. The "immigrant share" is the ratio of the number of émigrés in each country to the number of active mathematicians. The "quality-adjusted share" is the ratio of the total number of citations received by the émigrés in 1984-1989 to the total number of citations produced by the country's mathematicians in that period. The connectivity index measures the publication link between mathematicians in a particular country and other countries; see the text for a definition.

Table X

Determinants of international differences in impact of Soviet supply shock

Sample/Variables:	Measure of overlap index used in first-stage regression		
	Correlation coefficient	Index of intensity	Index of similarity
Ten largest countries:			
Immigrant share (%)	-0.076	-0.041	-0.382
	(0.029)	(0.022)	(0.130)
Connectivity index	-2.431	-1.812	-8.259
	(0.633)	(0.537)	(10.188)
R^2	0.806	0.744	0.814
All 26 countries:			
Immigrant share (%)	-0.013	-0.009	-0.062
	(0.007)	(0.005)	(0.037)
Connectivity index	-2.609	-1.620	-10.553
	(0.603)	(0.517)	(10.672)
R^2	0.230	0.167	0.374

Notes: Robust standard errors in parentheses. The dependent variable is the coefficient of the interaction between the index of overlap and the post-1992 dummy variable from the first-stage regression estimated separately in each country. The regressions in the top panel use the sample of 10 countries where there are at least 2,000 active mathematicians, while the regressions in the bottom panel use the 26 countries with at least 500 active mathematicians. The regressions are weighted by the inverse of the sampling variance of the dependent variable. The "immigrant share" is quality-adjusted, giving the ratio of the number of AMS citations received in 1984-1989 by Soviet émigrés who eventually moved to country i to the total number of citations received by mathematicians in country i over the same period. The connectivity index measures the publication link between mathematicians in a particular country and other countries; see the text for a definition.