

# WHICH PEERS MATTER? THE RELATIVE IMPACTS OF COLLABORATORS, COLLEAGUES, AND COMPETITORS

George J. Borjas and Kirk B. Doran\*

*Abstract*—Many economists believe knowledge production generates positive spillovers among knowledge producers. The available evidence, however, is mixed. We argue that spillovers can exist along three dimensions: idea, geographic, and collaboration space. To isolate the key channel through which knowledge spills over, we use a unique data set to examine the impact of a large post-1992 exodus of Soviet mathematicians on the output of the nonémigrés. Although the data reveal strong competitive effects in idea space, there is evidence of knowledge spillovers in collaboration space, when high-quality researchers directly engage with other researchers in the joint production of new knowledge.

## I. Introduction

**K**NOWLEDGE production has the potential to be the engine of economic growth. A key reason is the possibility that one knowledge producer may have strong positive spillovers on other knowledge producers (Lucas, 1998, 2009). These potential spillovers could offset the traditional law of diminishing returns, allowing for a cycle of long-term economic growth (Romer, 1986, 1990). One crucial barrier to progress is the fact that economists do not yet have a good empirical understanding of the shape of the knowledge production function and of the channels that allow the knowledge spillovers to occur (Jones, 2005).

A number of recent empirical studies exploit supply shocks generated by a variety of natural experiments to document the existence and measure the magnitude of such spillovers. For example, Waldinger (2010) examines the output of the doctoral students who were left behind when superstar German scientists emigrated during the Nazi era. He finds that these students suffered in the absence of their highly skilled mentors. In more recent work, however, Waldinger (2012) reports that the fellow professors left behind in the affected German university departments did not experience a loss in output when their colleagues left. In contrast, Azoulay, Graff Zivin, and Wang (2010) document the decreased productivity suffered by the coauthors of superstar scientists in biological labs after the superstars die unexpectedly. Finally, Borjas and Doran (2012) examine the impact of the large influx of Soviet mathematicians into the United States after the collapse of the Soviet Union and find that the output of American mathematicians with the most Soviet-like research programs fell dramatically.

The rapidly growing literature therefore reports mixed evidence on the hypothesis that human capital externalities play an important role in the knowledge production func-

tion. However, there are three conceptually distinct spaces in which knowledge producers live and work: the space of ideas, geographic location, and the collaboration network. The conflicting evidence in the empirical literature may be simply reflecting the possibility that the different natural experiments are measuring the spillovers resulting from supply shocks in different spaces.

In this paper, we make use of a single natural experiment to evaluate the importance of these conceptually distinct types of closeness. Specifically, the study examines the consequences of the mass emigration of Soviet mathematicians following the collapse of the Soviet Union on the output of the mathematicians who remained behind.<sup>1</sup> During much of the seventy-year rule of the Communist Party in the USSR, Soviet scientists faced severe restrictions on emigration and on communications and collaborations with Western scientists. After the political process that led to the collapse of the Soviet Union accelerated in 1990, Soviet scientists embraced the opportunity to consider job opportunities provided by Western institutions and began to migrate in large numbers to Western countries.<sup>2</sup> Nearly 10% of the active Soviet mathematics workforce moved to a country outside the borders of the former Soviet Union after 1992.

As a result of the diaspora, some Soviet mathematicians lost peers who were close to them in idea space (i.e., working on the same topics), other mathematicians lost peers who were close to them in geographic space (i.e., worked in the same university department), and still others lost peers who were close to them in collaboration space (i.e., they had been coauthors prior to the collapse). Our theoretical model suggests that a supply shock in any particular space will typically generate both spillovers and competition effects. The law of diminishing returns implies that the mass emigration should improve economic opportunities for the mathematicians left behind, while the spillover hypothesis implies that the emigration would reduce their future output. The observed effect represents the net impact of these two forces. Clearly, these forces could vary in strength depending on the space where the supply shock occurred.

The nature of the supply shock experienced by a particular Soviet mathematician in each of these spaces obviously depends on how many of the mathematician's peers in the particular space left the Soviet Union. In turn, the composi-

Received for publication May 30, 2013. Revision accepted for publication March 13, 2014. Editor: Philippe Aghion.

\* Borjas: Harvard University; Doran: University of Notre Dame.

We are grateful to Patrick Ion, Graeme Fairweather, Norm Richert, and Erol Ozil from the American Mathematical Society for extensive collaboration and support in preparing the data.

A supplemental appendix is available online at [http://www.mitpressjournals.org/doi/suppl/10.1162/REST\\_a\\_00472](http://www.mitpressjournals.org/doi/suppl/10.1162/REST_a_00472).

<sup>1</sup> Borjas and Doran (2012) summarize the history of mathematics in the Soviet Union between 1922 and 1992; see also Abramitzky and Sin (2014) and Ganguli (2012).

<sup>2</sup> The diaspora was also spurred by the disarray in the funding of Soviet scientific research that accompanied the collapse of the Soviet Union. To stem the exodus, George Soros established a program that provided research funds to scientists who chose to remain. See Ganguli (forthcoming) for an analysis of the impact of this program on career choices.

tion of the émigré subsample depends on which peers were most eager to emigrate.

Differences in the incentives to emigrate arose partly from the fact that (because of historical anti-Semitism and its consequences) Jewish mathematicians had a stronger motivation to leave the collapsing Soviet Union than did non-Jewish mathematicians. Because the collapse of the Soviet Union was unanticipated, mathematicians in the 1980s who chose locations in idea space, geographical space, or collaboration space with many high-quality Jewish peers could not have anticipated the fact that many of those peers would suddenly migrate in the early 1990s. We therefore use the relative number of a mathematician’s peers who have a Jewish identity in idea space, in their institution, or among their coauthors as instruments for the supply shocks in idea space, geographical space, and collaboration space, respectively.

Our empirical analysis suggests that competition effects dominate in the space of ideas: a 10 percentage point increase in the out-migration rate in idea space increased the number of papers that a Soviet nonémigré published in any given year by about 7%. At the same time, there is no consistent evidence of a link between a mathematician’s productivity and supply shocks in either geographic space or collaboration space for the average Soviet mathematician. However, our empirical analysis unveils the key setting in which spillovers strongly dominate competition effects: the loss of collaborators of very high quality. Ranking authors by the quality of their collaborators, we find that authors in the top 5% of this ranking suffer an 8% decline in publications for every 10% of their collaborators who emigrate.

In sum, our study synthesizes the diverse set of results found in the literature and advances a plausible explanation: the relative importance of human capital spillovers depends on the nature of the space where the supply shock occurs *and* on the quality of the individuals in question. Competition is the dominant consequence of interactions in the space of ideas and is important enough in collaboration space to render the loss of the average collaborator innocuous. But spillovers can dominate in the special case where a person of very high quality directly interacts with others in the joint creation of knowledge.

## II. Theory

Modern theories of economic growth emphasize how human capital externalities alter the productivity of specific workers. A worker surrounded by many high-skill workers will become more productive by being exposed to new ideas and concepts. These spillover effects, however, must coexist with the traditional laws of scarcity and diminishing returns that form the core of economic analysis.<sup>3</sup> Following

Jones and Romer (2010), suppose that the production function for “mathematical knowledge”  $Y$  in the Soviet Union depends on the stock of ideas  $A$ , the stock of resources  $K$  used as inputs (e.g., faculty slots and computing resources), and the stock of mathematicians  $L$ . The generic human capital spillovers model assumes constant returns to  $K$  and  $L$ . Suppose that the aggregate production function can be written as

$$Y = A^\phi (\alpha_K K^\delta + \alpha_L L^\delta)^{1/\delta}, \quad (1)$$

where  $\phi$  gives the “externalities elasticity,” the percent expansion in mathematical knowledge associated with a 1% increase in the stock of ideas; and the elasticity of substitution between labor and capital  $\sigma_{KL} = 1/(1 - \delta)$ .

It is typically assumed that the stock of ideas is proportional to the number of mathematicians (for simplicity,  $A = L$ ). A mathematician’s marginal product is then given by

$$MP_L = AP_L (\phi + \alpha_L Q^{-\delta} L^\delta), \quad (2)$$

where  $AP_L = Y/L$  and  $Q = [\alpha_K K^\delta + \alpha_L L^\delta]^{1/\delta}$ .

This model can be used to address a crucial question: What happens to the marginal product of mathematicians following a supply shock that simultaneously changes the number of ideas *and* the number of workers? It is easy to show that

$$d \log MP_L = \phi m + s_K \left( \frac{\phi + (1 - \delta) s_L}{\phi + s_L} \right) (d \log K - m), \quad (3)$$

where  $m = d \log L$ ;  $s_L = \alpha_L L^\delta / Q^\delta$ , or labor’s share of output; and  $s_K = 1 - s_L$ .<sup>4</sup> For expositional convenience, consider the case where  $m$  is positive. Equation (3) summarizes the opposing forces on the marginal product of mathematicians arising from the supply shock. On the one hand, the expansion in the stock of ideas makes the preexisting mathematicians more productive. On the other hand, the law of diminishing returns comes into play and reduces productivity as a result of the increased number of mathematicians.

It is instructive to illustrate the implications of the model in terms of two alternative scenarios: the short run and the long run. By definition, resources  $K$  are fixed in the short run and are fully adjusted in the long run. If the academic market were competitive, additional resources would enter or exit the market until the rate of return to these resources is again equal to the world rate. In a competitive market, the proportionate change in the capital stock would equal the proportionate change in the number of mathematicians

<sup>3</sup> Borjas and Doran (2015) use a simpler version of this model to examine the determinants of mobility across fields in mathematics.

<sup>4</sup> Specifically, the mathematician’s wage  $w$  equals the value of marginal product when the employer ignores the externality, or  $\alpha_L Q^{1-\delta} L^{\phi+\delta-1}$ . It follows directly that  $s_L = wLY$ .

(or  $m$ ). The resulting change in the marginal product of mathematicians would then be given by

$$d \log MP_L = \begin{cases} \left( (\phi - s_k) + \frac{\delta s_L s_k}{\phi + s_L} \right) m, & \text{if } d \log K = 0 \\ \phi m, & \text{if } d \log K = m. \end{cases} \quad (4)$$

Suppose  $Q$  is Cobb-Douglas, so that  $\delta$  is 0. Equation (4) then collapses to  $(\phi - s_k) m$  in the short run. Resources are fixed, and there is a race between the externalities elasticity and the law of diminishing returns. If spillovers were non-existent or small, the supply shock would have the usual effect of reducing the productivity of mathematicians. If the elasticity  $\phi$  is sufficiently large, the direction of the effects would reverse: the marginal product of mathematicians rises because of the expansion in the stock of ideas.<sup>5</sup>

Over time, the capital stock adjusts to the supply shock. In other words, there is a change in the number of faculty slots allocated to existing mathematics departments and journals change the number of pages published. If the institutions hiring mathematicians act as if they were competitive firms, the long-run adjustment in the capital stock would be given by  $d \log K = m$ . Equation (4) shows that the long-run productivity impact of a supply shock equals  $\phi m$ . In the absence of human capital externalities, the long-run impact of a supply shock on the productivity of mathematicians is 0. If spillover effects exist, however, the long-run impact is positive because the stock of ideas expanded.

#### A. Fields and Networks

The study of mathematics, like all other academic disciplines, is composed of many fields, and the theorems and techniques in the various fields join together to form the current stock of “mathematical knowledge.” Because the supply shock induced by the collapse of the Soviet Union had a differential impact on different fields, it is important to examine how the variation in the size of the supply shock across fields alters the lessons from the basic model.

The existence of distinct mathematical fields also allows us to generalize the notion of human capital externalities. Up to this point, our model has focused on the productivity impact of global human capital externalities. The specification of the aggregate production function in equation (1) implies that the change in the total stock of mathematical ideas shifts the productivity of all mathematicians by the same proportion. It is possible, however, that a mathematician’s productivity also depends on local externalities. For

example, it seems plausible that a supply shock that differentially affected the size of the workforce in, say, geometric topology and quantum theory would have differential impacts on the stock of ideas useful to researchers in those two areas. Similarly, a supply shock affecting a particular university may have strong repercussions on the productivity of mathematicians at that location but weaker repercussions on the mathematicians affiliated with geographically remote institutions. Finally, the externalities may be particularly strong among mathematicians belonging to the same collaboration network.

The simplest way to introduce local externalities is to interpret  $L$  in equation (1) as the number of efficiency units supplied by the entire mathematics workforce. Suppose there are  $N$  distinct mathematical fields or, more generally, networks. It is analytically convenient to use a nested CES framework to aggregate the contribution made by mathematicians in the various networks. The total number of efficiency units provided by the entire mathematics workforce is

$$L = \left( A_1^\theta L_1^\beta + \dots + A_N^\theta L_N^\beta \right)^{1/\beta}, \quad (5)$$

where  $A_n$  gives the stock of ideas in network  $n$ ,  $L_n$  gives the size of the mathematics workforce in field  $n$ ,  $\theta$  is the local externality elasticity measuring the size of network-specific spillovers, and the elasticity of substitution between mathematicians working in different networks is defined by  $\sigma_N = 1/(1 - \beta)$ .

Suppose that the aggregate production function in equation (1) is Cobb-Douglas ( $Y = A^\phi K^{1-\alpha} L^\alpha$ ) and that the stock of ideas in a network is proportional to the size of the workforce in that network. It is easy to show that the change in marginal product for mathematician  $i$  in network  $n$  resulting from a supply shock that potentially affects the number of mathematicians in all networks is

$$d \log MP_{in} = (1 - \alpha) d \log K + (\phi + \alpha - \beta) m + (\theta + \beta - 1) m_n, \quad (6)$$

where  $m = d \log L$  and  $m_n = d \log L_n$ .

The short- and long-run productivity impacts are then given by

$$d \log MP_{in} = \begin{cases} \left( \phi - s_k + \frac{1}{\sigma_N} \right) m + \left( \theta - \frac{1}{\sigma_N} \right) m_n, & \text{if } d \log K = 0 \\ \left( \phi + \frac{1}{\sigma_N} \right) m + \left( \theta - \frac{1}{\sigma_N} \right) m_n, & \text{if } d \log K = m. \end{cases} \quad (7)$$

The productivity impact of the generalized supply shock on a particular mathematician now depends on two distinct measures of the change in supply: the percent change in the total supply of mathematical efficiency units ( $m$ ) and the percent change in the number of mathematicians conduct-

<sup>5</sup> It is difficult to imagine that “capital” and mathematicians are easily substitutable in the production of new theorems. Equation (4) implies that if the elasticity of substitution between labor and capital is less than 1, the negative impact introduced by nonrival labor and resources gets larger, making it even more likely that the marginal product of mathematicians will fall in the short run following an increase in supply.

ing research in network  $n$  ( $mn$ ). In general, the short-run impact of these distinct measures of the supply shock can be either positive or negative, depending on the relative strength of the relevant externalities elasticity and the elasticities of substitution among the various groups. In the long run, the generalized supply shock  $m$  will have a positive impact (since  $m$  is assumed to be positive), but there will be distributional effects that may harm the mathematicians in networks that experienced the largest supply shocks.

By construction, the model presented above constrains mathematician  $i$  to belong to only one of these networks. It is straightforward to generalize the model by allowing additional levels in the CES nesting. For example, think of a specific network as composed of mathematicians in a particular mathematical field (e.g., partial differential equations). It is likely that some of the mathematicians in that field are frequent collaborators, while others have more distant interactions. By defining yet another level in the CES nesting, it is easy to model the impact of supply shocks at each level of the nesting: the generalized supply shock ( $m$ ), the field-specific supply shock, and the collaboration-specific supply shock. The analog of equation (7) would then contain a term for each of the relevant supply shocks that would affect the productivity of a particular mathematician, and each of the coefficients would measure the difference between the relevant externalities elasticity and the relevant elasticity of substitution.<sup>6</sup>

This framework has interesting implications for the interpretation of the findings reported in the flurry of recent studies that measure the quantitative importance of human capital externalities. The model implies that because these empirical studies exploit conceptually different supply shocks, they are, in fact, estimating conceptually different (net) spillover elasticities. Some studies estimate the spillover effects within a collaboration network (Waldinger, 2010; Azoulay et al., 2010). In other cases, the spillover effects are being measured within a geographic area (Waldinger, 2012) or at the field level (Borjas & Doran, 2012). Moreover, none of these exercises estimate the global externality  $\phi$  since all workers within a particular natural experiment are exposed to the same (total) shock, and hence the global externality elasticity cannot be uniquely identified.

In short, the theoretical framework highlights the importance of properly defining the nature of the “neighborhood,” or space in which a human capital spillover is going to be measured. Different natural experiments inevitably create supply shocks in different neighborhoods and allow us to

estimate different (net) externality elasticities. One lesson from the existing evidence is that it seems easiest to observe positive spillovers when the supply shock hits a “local” neighborhood, such as a high-quality collaboration network. Once the analysis moves to a larger neighborhood, the traditional laws of scarcity reappear and begin to dominate the data.

### III. Data

Our analysis examines how the output of mathematicians who remained in the Soviet Union was affected by the supply shock that followed the collapse of the USSR. We use a database provided to us by the American Mathematical Society (AMS) that reports the number of papers published by every mathematician in the world, by field and year, since 1939. The AMS data report the mathematician’s institutional affiliation at the time the paper was published, as well as the location of the affiliation, but this information is available on a systematic basis only since 1984. The AMS database also contains information on the number of citations the papers received, but the AMS counts only citations in a limited number of journals and reports only the post-2000 citations received by a paper (regardless of when the paper was published). Finally, the AMS professional staff assigns each publication 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 every mathematician in each of 63 different fields.

Borjas and Doran (2012) used these data, combined with information from the Thomson Reuters Institute for Scientific Information (ISI) Web of Science archive and the Mathematical Genealogy Project (MGP), to construct the universe of active Soviet mathematicians prior to the collapse of the Soviet Union and to define the subsample of 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. We then determined if the mathematician reported a Soviet affiliation in papers published before 1990 and classified anyone who used a Soviet affiliation at least half the time during this period as a “Soviet mathematician.” From this population, we defined 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.<sup>7</sup> The complement of this sample forms the population of Soviet “stayers”: Soviet mathematicians who remained within the geographic boundaries of the former Soviet Union after 1992. The online supplemental appendix contains a detailed description of the data and the construction of the samples.

These definitions yield a preexisting population of 12,224 active Soviet mathematicians. We also find that 1,051 of these Soviet mathematicians emigrated after the collapse of the Soviet Union. Table 1 reports summary sta-

<sup>6</sup> Consider a three-level nesting that defines the total labor supply of mathematicians ( $L$ ) in terms of the contribution of the supply in different networks,  $L_n$ , and then defines  $L_n$  in terms of the contribution of collaborators ( $L_C$ ) and noncollaborators in that network. It is easy to show that the short-run impact on the marginal product of mathematician  $i$  is given by

$$d \log MP_{mc} = \left( \phi - s_k + \frac{1}{\sigma_N} \right) m + \left( \theta - \frac{1}{\sigma_N} + \frac{1}{\sigma_C} \right) m_n + \left( \lambda - \frac{1}{\sigma_C} \right) m_c,$$

where  $\lambda$  is the collaboration-specific externalities elasticity and  $\sigma_C$  is the elasticity of substitution between collaborators and noncollaborators.

<sup>7</sup> By construction, an émigré must have published after 1992 (otherwise, we would never observe his post-1992 affiliation).

TABLE 1.—SUMMARY STATISTICS

	Soviet Émigrés	Soviet Nonémigrés
Number of papers		
1982–1991		
Mean	12.8	6.3
Median	10.0	4.0
Maximum	90.0	153.0
1992–2008		
Mean	28.3	7.6
Median	21.0	1.0
Maximum	317.0	311.0
Number of citations		
1982–1991		
Mean	38.5	6.7
Median	6.0	0.0
Maximum	1,256.0	2,240.0
1992–2008		
Mean	131.7	13.4
Median	44.0	0.0
Maximum	1,988.0	1,287.0
Number of mathematicians	1,051	11,173

The sample consists of active Soviet mathematicians who published at least one paper between 1970 and 1989 and were 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 Soviet Union.

tistics for the samples of Soviet mathematicians, both the émigrés and the nonémigrés. The most striking finding is the very large degree of positive selection that characterizes the Soviet émigrés. Prior to their migration, the émigrés were significantly more productive in terms of publications and citations. In particular, before 1992 the average (future) émigré had published 7 more papers and received 32 more citations than the typical Soviet who did not emigrate. After 1992, the émigrés' productivity far surpassed that of the Soviets who remained in the former USSR.<sup>8</sup>

The strong positive selection that characterizes the sample of émigrés suggests that the empirical analysis must carefully consider how the endogeneity of the out-migration decision influences the impact of the supply shock on the output of the mathematicians who remained in the former Soviet Union. We address the endogeneity issue in much greater detail below.

The data allow us to calculate the supply shock faced by a given Soviet mathematician in each of the three spaces defined earlier: the space of ideas, the space of geography, and the collaboration network.<sup>9</sup>

<sup>8</sup> Franzoni, Scellato, and Stephan (2012) and Grogger and Hanson (2013) report additional evidence on the positive selection of scientific migrants, particularly those who move to the United States.

<sup>9</sup> While geographic space is an obvious metric in which the idea of closeness can be easily grasped and closeness in the collaboration network can be understood intuitively in terms of the number of nodes separating two authors, the concept of closeness in idea space is more difficult to visualize. However, given the technology of proof production, there is a well-defined sense in which mathematicians who work on topics that build on each other are closer in the space of ideas than are mathematicians who work on disparate topics. When a mathematician works on a new theorem, he is usually inspired by intellectual work that has already been done in a similar area and then combines that work with new insights currently circulating in related fields of inquiry. In fact, recent research has developed ways of measuring the distance among fields of mathematics based on, for example, the frequency of copublishing across them (Basu & Dobler, 2012).

We use the field composition of mathematician  $i$ 's pre-1990 publication record to calculate the out-migration rate of Soviet mathematicians who had a similar field mix. In particular, let  $P_{f0}$  be the number of pre-1990 papers published in field  $f$  by mathematicians who eventually remained in the former Soviet Union, and let  $P_{f1}$  be the respective number of papers published by mathematicians who eventually emigrated. Let  $\alpha_{if}$  be the share of pre-1990 papers that mathematician  $i$  published in field  $f$ . The out-migration rate of persons who had a similar field mix as mathematician  $i$  is then given by

$$q_{iD} = \sum_f \alpha_{if} \frac{P_{f1}}{P_{f0} + P_{f1}}. \quad (8)$$

The variable  $q_{iD}$  gives the fraction of intellectually similar mathematicians who emigrated after the collapse of the Soviet Union, measuring the size of the supply shock experienced by mathematician  $i$  in idea space.

A second type of supply shock occurs at the geographic level. Some Soviet institutions may have been particularly hard hit by the outflow of mathematicians. Let  $k$  be the last institution employing mathematician  $i$  before 1990. Let  $P_{k0}$  be the number of pre-1990 papers published by mathematicians who chose to remain in the Soviet Union and whose last affiliation before 1990 was at institution  $k$ , and let  $P_{k1}$  be the respective number of papers published by mathematicians who eventually emigrated. The institution-specific out-migration rate affecting mathematician  $i$ , which measures the size of the supply shock in geographic space, can then be defined as

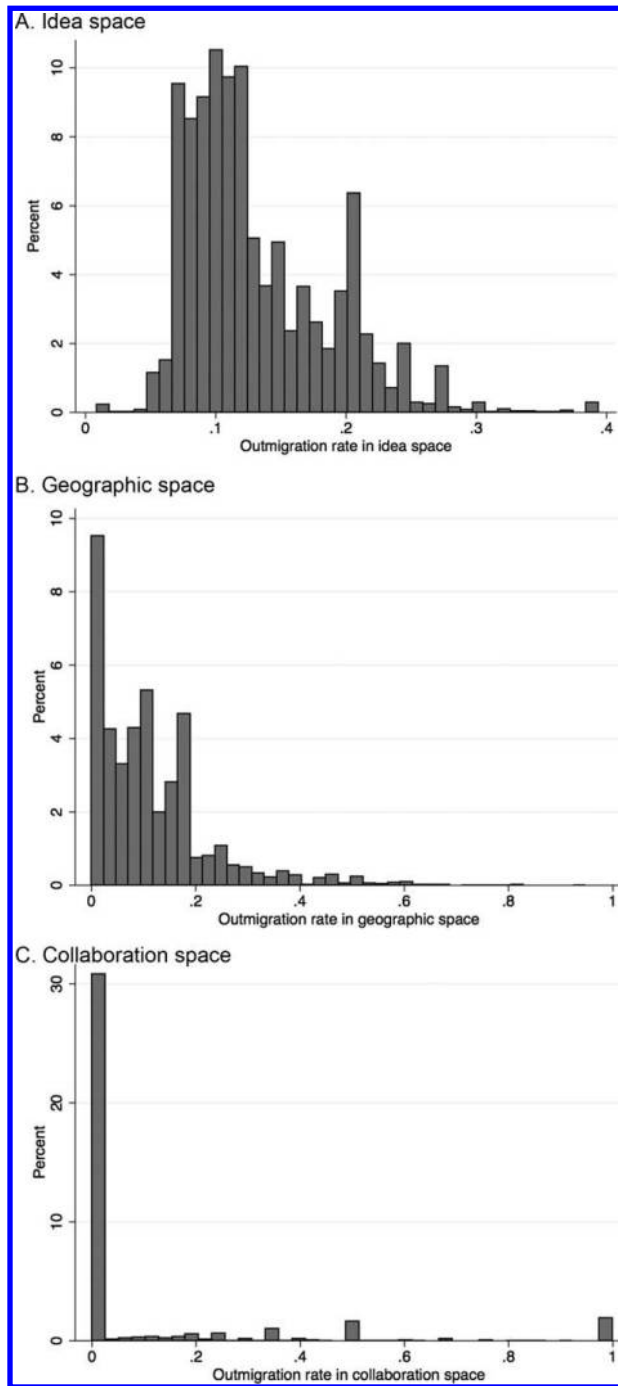
$$q_{iG} = \frac{P_{k1}}{P_{k0} + P_{k1}}. \quad (9)$$

Finally, mathematician  $i$  may have a network of collaborators  $N$ , some of whom remained in the Soviet Union after the collapse and some of whom left. The AMS provided us with a data set that details all coauthorship relationships in the universe of the 12,224 Soviet mathematicians. We can use the information on these collaboration networks to estimate the collaboration-specific supply shock. In particular, let  $P_{iC0}$  be the number of pre-1990 papers written by the collaborators of mathematician  $i$  who chose to remain in the Soviet Union, and let  $P_{iC1}$  be the respective number of papers written by the collaborators who emigrated. The collaboration-specific out-migration rate, which measures the supply shock in collaboration space, is then given by

$$q_{iC} = \frac{P_{iC1}}{P_{iC0} + P_{iC1}}. \quad (10)$$

The three panels of figure 1 illustrate the frequency distribution of the three supply shocks affecting the sample of mathematicians who remained in the former Soviet Union. All of the panels show a sizable dispersion in the supply

FIGURE 1.—FREQUENCY DISTRIBUTION OF OUTMIGRATION RATES



The out-migration rate measures the fraction of the peers along a particular dimension who emigrated from the Soviet Union after 1992. See the text for a precise definition.

shocks experienced by individual mathematicians. For example, the frequency distribution for the out-migration rate in idea space has a median of 11.7%, with the 10th and 90th percentiles being 7.3% and 20.8%, respectively. The mode of the supply shock in geographic space is 0.0%, indicating that over 20% of Soviet mathematicians were associated with departments that did not have any émigrés. Nevertheless, 25% of mathematicians were associated with

institutions that had out-migration rates exceeding 16.7%. Finally, the supply shock in collaboration space shows that about 20% percent of Soviet mathematicians lost at least one collaborator to the post-1992 diaspora.

Our empirical analysis examines the impact of these supply shocks on the output of the mathematicians who remained in the Soviet Union. We use the regression model

$$y_{it} = \varphi_i + \varphi_t + X_{it}\gamma + (\mathbf{q}_i \times T)\boldsymbol{\pi} + \varepsilon_{it}, \quad (11)$$

where  $y_{it}$  measures the output of mathematician  $i$  in year  $t$ ,  $\varphi_i$  is a vector of individual fixed effects,  $\varphi_t$  is a vector of year fixed effects,  $X_{it}$  is a vector of standardizing variables,  $\mathbf{q}_i$  gives the vector of person-specific supply shocks ( $q_{iD}$ ,  $q_{iG}$ ,  $q_{iC}$ ), and  $T$  is a dummy variable indicating if the observation is observed after 1992. The vector of variables  $X$  gives the years of work experience for mathematician  $i$ , defined as the number of years that have elapsed since the mathematician's first publication, and is introduced as a fourth-order polynomial.

This regression model uses a difference-in-differences approach. Many things happened simultaneously in the Soviet Union during the end of Communist party rule, which is why we compare mathematicians who did lose peers with mathematicians who did not, using the latter group as a control for all of the other changes that occurred at that time. We estimate the regression model using annual data between 1982 and 2008 and use two measures of a mathematician's output as dependent variables: an indicator variable set to unity if mathematician  $i$  published at all in year  $t$  and the number of papers published in that year. The theory suggests that the elements of the vector  $\boldsymbol{\pi}$  estimate the net spillover effect resulting from each particular type of supply shock.

The various columns of table 2 report the OLS coefficients from alternative specifications of the regression model in equation (11). It is evident that there is a strong positive correlation between each individual measure of the supply shock and the output of the Soviet mathematicians who remained behind. For example, a 10% increase in the out-migration rate in idea space increases the number of papers published annually by 0.09, a 10% increase in the out-migration rate in geographic space increases the number of papers published by 0.05, and the loss of all collaborators increases the number of papers by 0.09. Obviously a substantial correlation between the unobserved determinants of a mathematician's productivity and the supply shocks faced by that mathematician would bias the OLS coefficients.

#### IV. Accounting for Endogeneity

The subsample of Soviet émigrés residing in a particular mathematician's idea, geographic, and collaboration spaces is not random. The productivity profile of mathematicians who lose many intellectual peers to emigration likely differs from the productivity profile of those who lose few peers.

TABLE 2.—IMPACT OF SUPPLY SHOCK ON SOVIET NONÉMIGRÉS, OLS COEFFICIENTS

Variables	Specification			
	(1)	(2)	(3)	(4)
A. Publication Propensity				
Supply shock in idea space	0.393 (0.044)	—	—	0.349 (0.044)
Supply shock in geographic space	—	0.197 (0.023)	—	0.176 (0.024)
Supply shock in collaboration space	—	—	0.019 (0.012)	−0.013 (0.012)
B. Number of papers				
Supply shock in idea space	0.889 (0.140)	—	—	0.786 (0.140)
Supply shock in geographic space	—	0.425 (0.067)	—	0.371 (0.069)
Supply shock in collaboration space	—	—	0.057 (0.031)	−0.012 (0.033)

Standard errors are reported in parentheses and are clustered at the individual level. The publication propensity dependent variable equals 1 if the mathematician published at least one paper in any given year. The supply shock in intellectual space gives the out-migration rate of persons who have a similar field mix; the supply shock in geographic space gives the out-migration rate in the institution that last employed the mathematician prior to 1990; the supply shock in collaboration space gives the fraction of a mathematician's coauthors who emigrated after 1992. The table reports the coefficient of the interaction between the respective supply shock and a post-1992 dummy variable. The regressions have 283,169 observations. All regressions include the mathematician's years of experience (introduced as a quartic polynomial), year fixed effects, and individual fixed effects.

Similarly, there are unobserved factors that allowed some mathematicians in a particular collaboration network to receive job offers from the West. Any type of intranetwork correlation would contaminate the OLS estimate of the impact of the supply shock on the productivity of the coauthors that remained in the Soviet Union. Hence we use three separate instruments for the supply shocks in each mathematician's idea, geographic, and collaboration spaces.

All three of our instruments rely on the Jewish identity of the peers.<sup>10</sup> Jewish mathematicians were more likely to emigrate from the Soviet Union than their non-Jewish counterparts for two reasons. First, it has been repeatedly documented that Soviet-era anti-Semitism affected the Jewish mathematical community, even at the highest levels (Saul, 1999; Graham, 1993). Second, Israel was eager to welcome Jewish immigrants, granting them and their families automatic citizenship.

To construct the instruments, we need to know which Soviet mathematicians were likely to have Jewish ancestry. We do not have access to an externally supplied list of all active Jewish scientists. However, we do have access to the surnames of the 8,500 well-known Russian Jewish individuals (all of whom were born before the collapse of the Soviet Union) listed in the 1995 *Jewish Encyclopedia of Russia (Rossiyskaya Evreiskaya Entsiklopediya)*. In addition, the AMS provided us with the full names of the 12,224 active Soviet mathematicians in our data. Because many Jewish family names have distinctive Hebrew origins (such as מרגוליס, or Margulis) or are rarely used by non-Jews in Russia, we can identify mathematicians who are likely to have a Jewish identity by matching the surnames of mathematicians in the AMS database with the list of Jewish surnames in the *Encyclopedia*. This matching tech-

nique identified 1,051 likely Jewish names out of the 12,224 active Soviet mathematicians, or 8.6% of the active Soviet mathematics workforce.<sup>11</sup>

A simple regression of émigré status on likely Jewish identity is revealing. Mathematicians with a Jewish identity were more than twice as likely to emigrate than their non-Jewish peers (17.4% versus 7.8%). We can further check the accuracy of our detection of likely Jewish identity by determining how many of the émigrés who moved to Israel are predicted to be Jewish. Israel's immigration laws suggest that practically all of the Soviet mathematicians moving there had a Jewish identity (although many may have had gentile family names due to marriage or other reasons). In fact, over 30% of the émigrés who moved to Israel had one of the Jewish surnames listed in the *Encyclopedia*.<sup>12</sup>

We construct the instrument for the out-migration rate in the space of ideas as follows. In particular, let  $P_{fS}^*$  be the total number of pre-1990 papers published in field  $f$  by all Soviet mathematicians and  $P_{fJ}^*$  be the respective number of pre-1990 papers published by Soviet Jewish mathematicians. We can then define an index measuring the proportion of mathematician  $i$ 's peers in idea space that are at risk of migrating at the time of the collapse as

$$q_{iD}^* = \sum_f \alpha_{if} \frac{P_{fJ}^*}{P_{fS}^*}, \quad (12)$$

<sup>11</sup> According to the Soviet Census of 1989, one-half of 1% of the population in the Soviet Union had Jewish ancestry (*Demoskop Weekly*, 2013).

<sup>12</sup> We tried several alternative approaches to determine likely Jewish identity based on by-hand inspection of the list of names of Soviet émigrés and by-hand Internet searches of individual names and even biographical information about individual mathematicians. Every approach we tried yielded remarkably similar results. We report the results that use the surname list in the *Jewish Encyclopedia of Russia* because they are the simplest to explain and replicate.

<sup>10</sup> This instrument has been used successfully in related studies; see Waldinger (2010, 2012, 2013) and Moser, Voena, and Waldinger (2014).

TABLE 3.—FIRST-STAGE REGRESSIONS

Instrument	Dependent Variable					
	Idea Supply Shock	Geographic Supply Shock	Collaborator Supply Shock	Idea Supply Shock	Geographic Supply Shock	Collaborator Supply Shock
	(1)	(2)	(3)	(4)	(5)	(6)
Idea ( $q_{iD}^*$ )	1.019 (0.023)	—	—	1.017 (0.023)	0.482 (0.043)	0.680 (0.082)
Geographic ( $q_{iG}^*$ )	—	0.299 (0.023)	—	-0.002 (0.005)	0.274 (0.022)	0.084 (0.022)
Collaborator ( $q_{iC}^*$ )	—	—	0.227 (0.021)	0.005 (0.003)	0.028 (0.006)	0.211 (0.021)
Angrist-Pischke $F$ -test of excluded instruments	1,979.28	175.71	121.99	1,428.53	131.43	88.89

Standard errors are reported in parentheses and are clustered at the individual level. The regressions have 283,169 observations. All regressions include the mathematician's years of experience (introduced as a fourth-order polynomial), a vector of year fixed effects, and a vector of individual fixed effects. See the text for a detailed description of the construction of the instruments.

where  $\alpha_{if}$  gives the share of pre-1990 papers that mathematician  $i$  published in field  $f$ .

There is little reason to believe that a Soviet mathematician's pre-1990 field mix was chosen with the knowledge that fields with Jewish mathematicians would suffer a sudden loss of talent after 1992. Borjas and Doran (2012) summarize historical evidence showing that the collapse of the Soviet Union was unexpected on both sides of the Iron Curtain.

We use an analogous approach to define the instrument for the supply shock in geographic space. Let  $k$  be the last institution employing mathematician  $i$  prior to 1990,  $P_{kJ}^*$  be the number of pre-1990 papers written by Jewish mathematicians associated with institution  $k$ , and  $P_{kS}^*$  be the respective number of papers written by all Soviet mathematicians in that institution. We define the instrument for the supply shock in geographic space as

$$q_{iG}^* = \frac{P_{kJ}^*}{P_{kS}^*}. \quad (13)$$

Note that the instrument in equation (13) has a value approaching 1 when all mathematicians affiliated with institution  $k$  have a Jewish identity and 0 when none do.

Finally, let  $\alpha_{ip}$  be an indicator variable set to unity if the mathematician's coauthor in collaboration event  $p$  ( $p = 1, \dots, P$ ) has a Jewish surname.<sup>13</sup> We then define the instrument as

$$q_{iC}^* = \frac{1}{P} \sum_p \alpha_{ip}, \quad (14)$$

where the summation is conducted over all collaborations that occurred prior to 1990. The instrument therefore is

<sup>13</sup> A collaboration event is a pairing between two authors in a single paper. If mathematician  $i$  writes a paper with mathematicians  $j$  and  $k$ , then mathematician  $i$  has two collaboration events in that paper (one with mathematician  $j$  and one with mathematician  $k$ ).

given by the (weighted) fraction of mathematician  $i$ 's collaborators who are predicted to have a Jewish identity.<sup>14</sup>

## V. Results

Table 3 presents the relevant coefficients from the first-stage regressions. The first three columns show the relation when each instrument is included separately in the relevant regression, and the last three columns show the coefficients when all three instruments are introduced simultaneously. The instruments are sufficiently uncorrelated that they are not collinear and can be included either separately or simultaneously as instruments for the three supply shock variables in idea, geographic, and collaboration space, respectively.<sup>15</sup> Each instrument has a positive and significant impact on the proportion of a mathematician's peers who emigrated (in the relevant space). The Angrist-Pischke multivariate  $F$  test of excluded instruments is always highly significant, with  $F$  statistics of 88 or higher and  $p$ -values extremely close to 0.

Table 4 presents various alternative IV specifications of the regression model in equation (11). The evidence clearly indicates that the supply shock in idea space has a positive and significant effect on a mathematician's productivity. A supply shock that removes 10% of the mathematicians from a particular location in idea space leads to an increase of 0.06 in the number of papers that a competing mathematician publishes in any given year (using the specification in column 4). Since the average nonémigré published 0.8 papers per year prior to 1992, the 10 percentage point increase in the out-migration rate is associated with an 8% increase in output. Notably, the coefficients of the supply shock in geographic space and the supply shock in collaboration space are frequently or always insignificantly different from 0.<sup>16</sup> It

<sup>14</sup> We also experimented with specifications where the fraction is not weighted; the results were quite similar to those reported below.

<sup>15</sup> The pairwise correlations among the instruments for the shocks in idea, geographic, and collaboration space range from about 0.10 to 0.15.

<sup>16</sup> We also experimented with alternative demand-side instruments that predict the migration of peers in idea and geographic space using information on the preshock worldwide demand for the topics the peers worked on. The results are similar to those reported in the text.



TABLE 4.—IMPACT OF SUPPLY SHOCK ON SOVIET NONÉMIGRÉS, IV COEFFICIENTS

Variables	Specification					
	(1)	(2)	(3)	(4)	Only Top Journals (5)	Coauthor Share (6)
A. Publication Propensity in a Given Year						
Supply shock in:						
Idea space	0.570 (0.088)	—	—	0.357 (0.105)	0.376 (0.048)	—
Geographic space	—	0.366 (0.080)	—	0.316 (0.095)	−0.023 (0.048)	—
Collaboration space	—	—	0.118 (0.058)	0.009 (0.067)	−0.026 (0.032)	—
B. Number of papers in a given year						
Supply shock in:						
Idea space	0.791 (0.250)	—	—	0.563 (0.294)	0.532 (0.073)	1.049 (0.203)
Geographic space	—	0.471 (0.310)	—	0.432 (0.365)	0.012 (0.069)	0.374 (0.219)
Collaboration space	—	—	0.092 (0.142)	−0.063 (0.173)	−0.015 (0.049)	0.103 (0.125)

Standard errors are reported in parentheses and are clustered at the individual level. The regressions have 283,169 observations. All regressions include the mathematician's years of experience (introduced as a fourth-order polynomial), a vector of year fixed effects, and a vector of individual fixed effects. See the text for a detailed description of the instruments.

therefore seems as if the various supply shocks sparked by the collapse of the Soviet Union had their most consistent average impact in idea space, where they significantly reduced competition and allowed the stayers to publish more.<sup>17</sup>

The raw data clearly reveal the source of the IV results reported in table 4. In particular, figure 2 illustrates the trends in the annual publication rate for the sample of mathematicians who remained in the former Soviet Union. In the first two panels, we identify two groups of mathematicians based on the value of each respective instrument: the first group consists of mathematicians who have values of the instrument below the 10th percentile, while the second group consists of those with values above the 90th percentile. As figure 2a shows, there is a relative increase after 1992 in the number of publications by mathematicians most exposed to a supply shock in idea space.

Figure 2b shows that there may have been a small relative increase after 1992 in the number of publications by mathematicians most exposed to a supply shock in geographic space, but the effect is smaller, less visually precise, and ended quickly, corresponding with the variation in the regression coefficients reported in table 4. The bottom panel of the figure, which illustrates the supply shock in collaboration space, defines the groups based on the presence of

Jewish coauthors. It is evident that there is little difference in the shape of the output profiles of mathematicians in the two groups.<sup>18</sup>

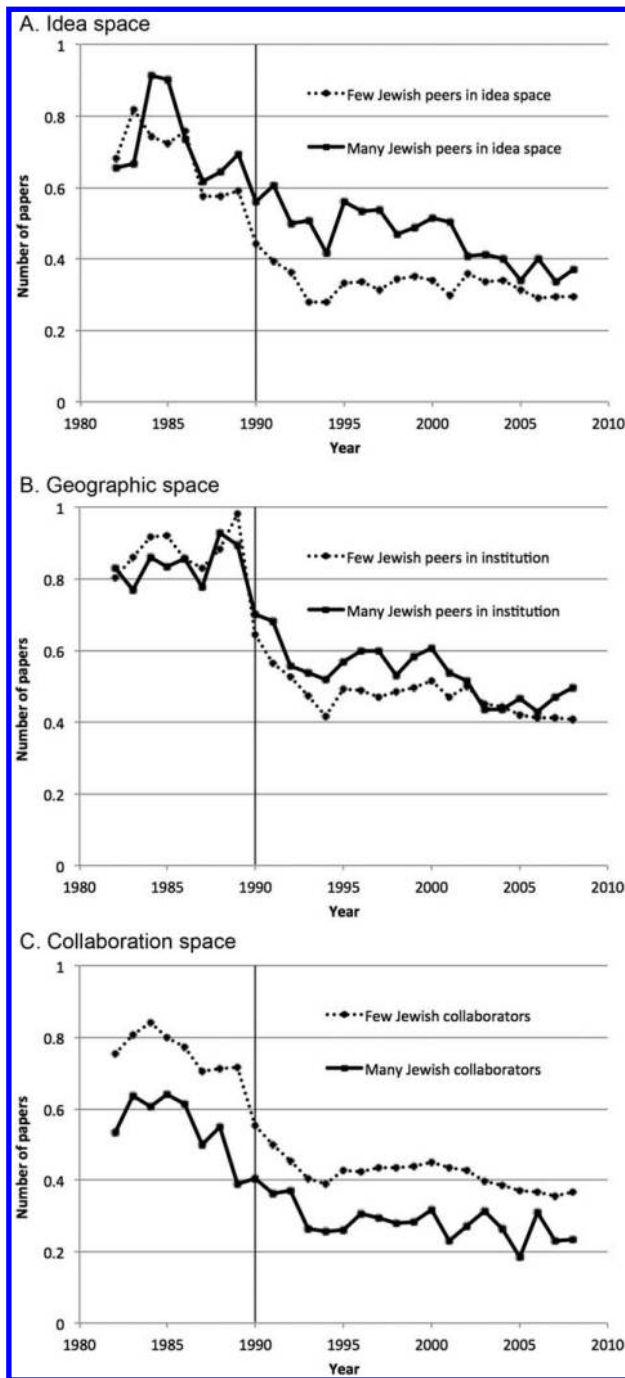
There are several reasons that Soviet stayers who lost many intellectual competitors in idea space may have published more. First, émigrés were less likely to continue publishing in Russian-language journals after departing; if the journals were resistant to decreasing their pages in subfields heavily hit by departures, this would have freed up space in the journal market for Soviet stayers in those fields. While such effects of the journal market on knowledge generation are important to document, it is also intriguing to consider another possibility: perhaps the departure of intellectual competitors gave the stayers access to jobs, research funds, graduate students, or other resources that improved their “true” productivity.

We examine this possibility by considering a market relatively unaffected by variation in Soviet journal space: the market for space in top international journals. We define top international journals as those with high-impact factors according to the Thompson Reuters Web of Science. Specifically, we focus on an extremely selective list of about forty exceptional journals across all areas of pure mathematics, all areas of applied mathematics, theoretical physics, statistical physics, and mathematical physics. The list includes outlets such as *Science*, *Nature*, the *Journal of the American Mathematical Society*, *Physical Review Letters*, and *Communications on Pure and Applied Mathematics* (see the online appendix for more details). We reestimated equation (11) using the number of papers published in these

<sup>17</sup> In addition to the main results using number of papers per year as the outcome variable, we also report ancillary results using publication propensity (any publications at all) as the outcome variable. Although the use of this binary outcome variable creates heteroskedasticity that could bias the standard errors, the Huber-White robust standard errors correct for this model-based heteroskedasticity in linear probability regressions, and the clustered standard errors that we report in the table do so as well. While logit regressions would avoid such model-based heteroskedasticity, they would introduce the very difficult numerical problem of maximizing a likelihood function with thousands of individual fixed effects. We follow much of the empirical literature and make use of linear regressions for these ancillary specifications, relying on Huber-White standard errors (and their cluster generalization) to correct for any resulting heteroskedasticity.

<sup>18</sup> Mathematicians who do not have a Jewish coauthor make up 79% of the sample, so we cannot divide the categories in figure 2C as we did in the first two panels. For definitional symmetry, we divide the categories according to no Jewish coauthors (the 79th percentile and below) and all Jewish coauthors (the 97th percentile and above). The results look similar if we expand the Jewish coauthors group to the 90th percentile and above (more than one-fourth Jewish coauthors).

FIGURE 2.—SCATTER OF PUBLICATIONS PER YEAR BY VALUE OF THE INSTRUMENTS, 1982–2008



The sample consists of Soviet mathematicians who remained in the Soviet Union after 1992.

top journals as the dependent variable. The regression coefficient for the supply shock in idea space is 0.11 (0.05), implying that a 10 percentage point increase in the out-migration rate is associated with a 22% increase in output in top journals.

One problem with emphasizing publications in the top international journals is that Soviet academics were strongly “discouraged” from publishing in Western outlets before the collapse of Soviet communism and needed special per-

mission to do so. In an environment in which even the best scientists were regularly arrested or banished (see the list of scientific prisoners in Graham, 1993) and in which any scientific contact with the West could result in punishment, the relative trickle of Soviet output in top international journals before 1992 represents something other than the sum total of high-quality Soviet achievement. Rather, it represents the small subset of high-level achievement that the authorities deemed acceptable for foreign publication. A preferable approach might be to compare output in the best outlets that Soviet academics could freely publish in before the collapse with output in the best outlets that Soviet academics could freely publish in after the collapse. To do so, we create a variable that is identical to the top 40 output measure defined earlier in the years after the collapse, but uses output only in the top three Soviet-era journals in the years before the collapse: *Математический сборник*; *Izvestiya Rossiiskoi Akademii Nauk, Seriya Matematicheskaya*; and *Uspekhi Matematicheskikh Nauk*.<sup>19</sup> We reestimate equation (11) using this alternative dependent variable and report the resulting coefficients in column 5 of table 4. We again find a large and statistically significant impact of the shock in idea space.<sup>20</sup>

The evidence therefore strongly suggests that Soviet scientists who lost intellectual competitors were indeed able to gain access to resources other than pages in (the now more empty) Russian-language journals. This finding is not surprising, considering that the fierce competition for limited research positions familiar in academia today was at least as prevalent in the scientific system of the Soviet Union (see Graham, 1993). Soviet scientists often waited years for someone in a research post to retire. The sudden migration of key scientists would have freed important research posts for the competitors who remained behind, especially in comparison to stayers, who lost few competitors during the shock but suffered from the same overall trends.

Up to this point our analysis gives each author of a paper full credit for the paper, regardless of the number of coauthors. If the incidence of coauthorship is changing over time and perhaps changing differentially across fields, the evidence reported in table 4 could be contaminated by our measurement decision. To examine the robustness of our results, we created a new outcome variable consisting of the total number of fractional shares of papers that an

<sup>19</sup> There was less international competition for top Soviet journals before the collapse than there was for top international journals after the collapse, so to hold quality constant, it is sensible to restrict the pre-collapse Soviet journals to a much smaller number. There are a small number of missing observations in the disaggregated data set provided to use by the AMS that contains full author and journal information. The results are almost identical if we treat missing observations as zeros in columns 5 and 6 of table 4 or if we simply exclude the observations. For simplicity, we report only the former set of results.

<sup>20</sup> We can also create a variable that uses output in both top Soviet and top international journals in the years before the collapse and only output in top international journals after the collapse. The resulting coefficient is similar in magnitude and still statistically significant.

TABLE 5.—IMPACT OF SUPPLY SHOCK, ADJUSTING FOR FIELD-YEAR TRENDS, IV COEFFICIENTS

Variables	Specification					
	(1)	(2)	(3)	(4)	Only Top Journals	Coauthor Share
					(5)	(6)
<b>A. Publication Propensity in a Given Field in a Given Year</b>						
Supply shock in						
Idea space	0.012 (0.002)	—	—	0.007 (0.002)	0.006 (0.001)	—
Geographic space	—	0.008 (0.002)	—	0.007 (0.003)	−0.000 (0.001)	—
Collaboration space	—	—	0.002 (0.001)	−0.000 (0.002)	−0.000 (0.001)	—
<b>B. Number of Papers in a Given Field in a Given Year</b>						
Supply shock in:						
Idea space	0.012 (0.004)	—	—	0.009 (0.005)	0.008 (0.001)	0.013 (0.003)
Geographic space	—	0.007 (0.005)	—	0.007 (0.006)	0.000 (0.001)	0.005 (0.003)
Collaboration space	—	—	0.002 (0.002)	−0.001 (0.003)	−0.000 (0.001)	0.002 (0.002)

Standard errors are reported in parentheses and are clustered at the individual level. The regressions have 18,122,816 observations. All regressions include the mathematician's years of experience (introduced as a fourth-order polynomial), a vector of year fixed effects, a vector of individual fixed effects, and a full vector of all field-year fixed effects. See the text for a detailed description of the instruments. Because the unit of observation used in the regressions is at the author-year-subject level, the coefficients need to be multiplied by 63 (for 63 fields of mathematics) to be comparable to those reported in this table.

author published in a given year, where an author earns a  $(1/n)$  share of a paper in which there are  $n$  authors. The results in column 6 again show a large and statistically significant impact of the shock.

Finally, one potential problem with the analysis is that once Soviet mathematicians were exposed to the Western mathematics community after 1992, the exposure may have increased the relative marketability of some types of mathematics. In particular, the sudden interaction with the demand for (and supply of) Western-style mathematics may affect the results by making it either easier (or harder) to publish in Western-style fields after the shock, creating a differential time trend in output across fields of mathematics. While our instrument is designed to avoid such issues (we have no reason to suspect that Soviet Jewish mathematicians chose to cluster in Western-style fields, nor is there a high correlation between the fields they specialized in and the fields that were popular in the West), we also address the problem by controlling for arbitrary field-specific secular trends in the regression model. The sample is now constructed so that a particular observation represents an author-field-year cell. In other words, our data now consist of the publication history of each Soviet nonémigré in each of the 63 fields in mathematics for each year in the sample period. The expanded regression model is

$$y_{ift} = \phi_i + \phi_{ft} + X_{it}\gamma + (q_i \times T)\pi + \varepsilon_{ift}, \quad (15)$$

where  $y_{ift}$  measures the productivity of mathematician  $i$  in field  $f$  in year  $t$  and  $\Phi_{ft}$  is a vector that includes all interactions between field and year fixed effects. Once again,  $\mathbf{q}_i$  gives the vector of person-specific supply shocks ( $q_{iD}$ ,  $q_{iG}$ ,  $q_{iC}$ ) defined earlier.

Table 5 reports the relevant coefficients from this expanded specification. A comparison of the coefficients in

tables 4 and 5 makes it evident that the results are qualitatively similar. The numerical magnitude of the coefficients is smaller in table 5, but this is purely a mechanical effect because the total impact of the supply shock is divided among the 63 fields that make up the mathematics discipline. The correct comparison is obtained by multiplying the coefficients in table 5 by 63. Such a comparison reveals that the impact of the supply shocks on the number of papers published is essentially the same in the two tables.<sup>21</sup>

The model presented in section II provides a useful lens for interpreting the regression evidence.<sup>22</sup> The impact of a specific type of supply shock depends on the relative strengths of human capital spillovers and the relaxation of the scarcity constraints. The positive effect of the supply shock in the space of ideas implies that diminishing returns play a very strong role in that particular space. In contrast, the precisely estimated 0 in collaboration space clearly indicates that the beneficial human capital spillovers are offset by the power of the diminishing returns.

<sup>21</sup> It is also important to consider the impact of selection bias: the mathematicians who remained in those areas of geographic, intellectual, and collaboration space who lost many peers may be intrinsically different from the mathematicians who remained in those areas who lost few peers. The inclusion of person-specific fixed effects controls for average differences in performance. It may be, however, that the slope of the productivity profile also differs between the two groups. These differences in slope, however, can be controlled for by estimating a regression that has individual linear time trends (i.e., interactions between the individual fixed effects and years of experience). We estimated various specifications of the regression model using this more general specification, and the results repeatedly confirmed the increase in the number of papers documented in tables 4 and 5.

<sup>22</sup> We do not have complete data on the number of citations received by the Soviet papers written prior to 1999. We note, however, that in related work where the citation data are more complete, the results tend to be similar regardless of whether the dependent variable is the number of papers or the number of citations (Borjas & Doran, 2012, 2015).

TABLE 6.—QUALITY OF COAUTHORS AND THE SUPPLY SHOCK IN COLLABORATION SPACE, IV COEFFICIENTS

	Specification			
	(1)	(2)	Coauthor Share (3)	Excludes Papers Coauthored with Émigrés (4)
<b>A. Publication Propensity</b>				
Supply shock in:				
Idea space	—	0.335 (0.106)	—	0.572 (0.108)
Geographic space	—	0.326 (0.095)	—	0.446 (0.099)
Collaboration space				
×Low-quality coauthors	0.162 (0.061)	0.045 (0.071)	—	0.074 (0.070)
×High-quality coauthors	−0.316 (0.091)	−0.355 (0.091)	—	−0.207 (0.098)
<b>B. Number of Papers</b>				
Supply shock in:				
Idea space	—	0.529 (0.295)	1.017 (0.205)	1.223 (0.281)
Geographic space	—	0.448 (0.365)	0.389 (0.220)	0.751 (0.305)
Collaboration space				
×Low-quality coauthors	0.160 (0.152)	−0.007 (0.184)	0.156 (0.133)	0.181 (0.174)
×High-quality coauthors	−0.583 (0.184)	−0.638 (0.190)	−0.427 (0.142)	−0.427 (0.179)

Standard errors are reported in parentheses and are clustered at the individual level. The regressions in columns 1–4 have 283,169 observations; the regressions in column 5 have 18,122,816 observations. All regressions include the mathematician's years of experience (introduced as a fourth-order polynomial), a vector of year fixed effects, and a vector of individual fixed effects. The regressions in column 5 also include a vector of field-year fixed effects. The "low-quality" and "high-quality" coauthor variables are indicator variables set to unity if the average quality of a mathematician's coauthors placed the author in the bottom 95% or top 5% of the quality distribution. See the text for a detailed description of the instruments.

At first glance, the result that a supply shock in collaboration space has no effect on productivity is surprising; Waldinger (2010) and Azoulay et al. (2010) each document that the loss of collaborators or mentors results in a substantial reduction in output among those who remain. However, both of these studies focus on supply shocks associated with the loss of extremely high-quality coauthors and mentors. This insight suggests that we estimate a separate set of regressions in which we allow the effect of the supply shock in collaboration space to vary depending on the quality of the coauthors.

We therefore rank authors according to the average quality of their coauthors. Table 6 reports the estimated coefficients from regressions where we interact the variable measuring the supply shock in collaboration space with two indicator variables set to unity if mathematician  $i$  is below or above the 95th percentile of this ranking.<sup>23</sup> The interacted variables show that the impact of the supply shock on output is consistently negative and significant for mathematicians with high-quality coauthors and typically insignificant for mathematicians with lower-quality coauthors. It is evident, therefore, that the effects of spillovers dominate the effects of competition in the subsample of mathemati-

cians who collaborate with very high-quality knowledge producers.<sup>24</sup> Moreover, the numerical impact of the supply shock in this select subsample is sizable. The loss of just 10% of a mathematician's high-quality coauthors reduces output per year by 0.06 papers (in the specification in column 2), implying roughly an 8% decline in publications.<sup>25</sup>

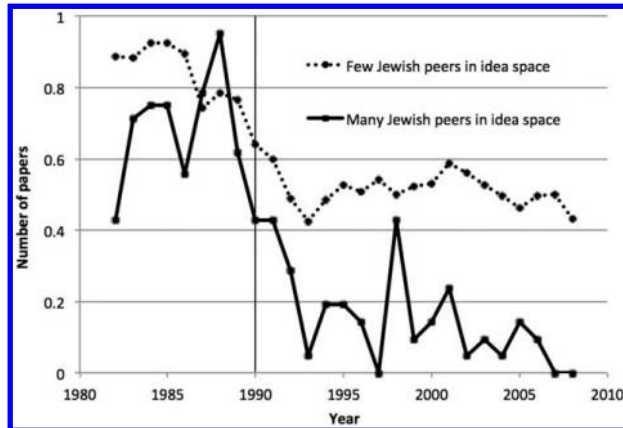
Figure 3 vividly illustrates this result by showing the trends in the annual publication rate for the subsample of mathematicians who remained in the former Soviet Union and were in the top 5% of the coauthor quality distribution. The figure shows the productivity path for two values of the instrument: the first group consists of persons who had no Jewish coauthors, and the second group consists of persons who had only Jewish coauthors. It is evident that mathematicians who had Jewish coauthors (who were much more likely to emigrate) had fairly similar levels of output before

<sup>24</sup> We can analogously allow the impact of the geographic shock to vary depending on the quality of the institutional colleagues. However, using a measure of colleague quality equivalent to that calculated for coauthors, we do not find any consistent evidence of statistically significant differences in the impact of the institutional shock at different levels of institutional quality. We cannot construct an analogous test in idea space: measuring field quality would require ranking the subfields of mathematics according to some absolute standard of achievement that transcends any one of them.

<sup>25</sup> The fact that the net beneficial spillovers are concentrated only around the highest-quality mathematicians and can be seen to accrue only to their actual coauthors could be taken to imply that spillovers in knowledge production are not generally important. However, the productivity impact of even these "limited" spillovers on some broad measure of mathematical knowledge could be quite substantial.

<sup>23</sup> We measure each coauthor's quality by calculating the percentile of his lifetime output in each field and then taking the paper-weighted average of these percentiles across all fields that he or she contributed to. We have tried a number of different ways of allowing the effects of loss of coauthors to vary according to coauthor quality (including linear interactions of coauthor quality and separate regressions for high-quality coauthors). The results are robust to specification changes.

FIGURE 3.—SCATTER OF PUBLICATIONS PER YEAR BY VALUE OF THE INSTRUMENT IN SUBSAMPLE OF SOVIET MATHEMATICIANS WITH HIGH-QUALITY COAUTHORS



The sample consists of Soviet mathematicians who remained in the Soviet Union after 1992 and had high-quality coauthors, defined as having coauthors where the average quality placed them in the top 92nd percentile of the quality distribution.

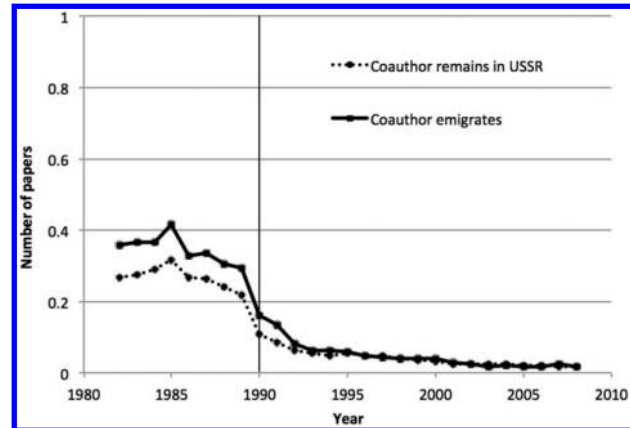
the shock to mathematicians with non-Jewish coauthors but much lower output after.

We can also use the data to determine how the departure of a coauthor from the Soviet Union adversely affected the productivity of the coauthor left behind. It is easy to document the break in the collaboration linkage resulting from the collapse of the Soviet Union. Figure 4 illustrates the trend in the average annual number of papers written by a coauthorship pair between 1982 and 2008 for two different types of collaborations: coauthorship pairs where both coauthors remained in the Soviet Union and coauthorship pairs where one coauthor emigrated. It is clear that there was a much greater decline in the intensity of collaboration in the latter type of coauthorship pairs. In short, the geographic separation dramatically increased the costs of collaboration between the émigrés and the nonémigrés.

As in Azoulay et al. (2010), we can document the impact of the shock in collaboration space on independent work by creating a new outcome variable: the number of papers that mathematician  $i$  wrote in year  $t$  that are not coauthored with future émigrés. We report the results using this outcome variable in column 4 of table 6. The impact of the supply shock in collaboration space remains negative and significant for those mathematicians with high-quality coauthors, and the numerical magnitude of the impact is roughly the same as that reported in the other specifications of table 6.

In sum, the coauthors who remained in the Soviet Union were adversely affected in two ways by the departure of their high-quality collaborators: they published less joint work with their now-absent coauthors, and they also published less of their own work as well. It seems as if the severing of the coauthorship relationship made it harder to pick up the knowledge that would have spilled over from the high-quality collaborator. For example, a mathematician will miss out on all the benefits that would have accrued from frequently discussing various topics with his collaborator or from testing out new ideas in casual conversa-

FIGURE 4.—BREAK IN COAUTHORSHIP LINKAGE FOLLOWING THE COLLAPSE OF THE SOVIET UNION, 1982–2008



The figure gives the average number of papers written in any given year by coauthorship pairs within each group.

tion. These types of informal activities play a key role in the transmission of tacit knowledge.

## VI. Summary

Many economists believe that knowledge production generates large positive externalities among knowledge producers, allowing for steadily increasing technological change and economic growth. We argue that these spillovers may occur in at least three distinct dimensions: spillovers may arise among knowledge producers who work on similar topics or ideas, or among knowledge producers who work in the same university department, or among knowledge producers who collaborate directly in the joint creation of knowledge.

We exploit the diaspora resulting from a single natural experiment to measure the productivity impact of a supply shock in each of these conceptually distinct types of spaces. In particular, we examine the consequences of the mass emigration of Soviet mathematicians after the collapse of the Soviet Union on the productivity of the mathematicians who remained behind. Some Soviet mathematicians lost peers who were close to them in idea space (i.e., they worked on the same topics), other mathematicians lost peers who were close in geographic space (i.e., they were employed by the same university department), and some lost peers who were close to them in collaboration space (i.e., they had been coauthors prior to the collapse).

Our theoretical model suggests that a supply shock in any particular space will typically generate both spillover and competition effects. The law of diminishing returns implies that the mass emigration should improve economic opportunities for the mathematicians left behind, while the spillover hypothesis would imply that the mass emigration should reduce their future output. The effect observed empirically represents the net impact of these two forces. Clearly, each of these forces may vary in intensity depending on the nature of the space being considered.

Our empirical analysis used a unique data set that contains information on the complete publication history of mathematicians who were active in the Soviet Union prior to its collapse. These data allow us to trace the impact of the diaspora on the output of Soviet mathematicians and to determine the relation between these trends and the supply shocks that occurred along the various dimensions. The evidence indicates that the supply shock in idea space led to an increase in academic output for the Soviet mathematicians left behind; the supply shock in geographic space had no consistent statistically significant effect on academic output; and the supply shock in collaboration space led to no change in research output for the average mathematician, but led to a large decline for mathematicians who lost high-quality coauthors in the exodus.

Our evidence emphasizes the fundamental but insufficiently studied fact that there are different ways of defining a peer. As a result, spillovers may be more or less powerful relative to competition for resources depending on which type of peer group is being examined. Competition is too intense in the space of ideas for spillovers to be empirically important among researchers who are not collaborating but are working on the same topics. Spillovers are more likely to be empirically relevant when two researchers are interacting regularly, are jointly producing new intellectual content, and at least one of them is of extremely high quality. Knowledge spillovers, in effect, are like halos over the heads of the highest-quality knowledge producers, reflecting only on those who work directly with the stars.

## REFERENCES

- Abramitzky, Ran, and Isabelle Sin, "Book Translations as Idea Flows: The Effects of the Collapse of Communism on the Diffusion of Knowledge," *Journal of the European Economic Association* 12 (2014), 1453–1520.
- Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang, "Superstar Extinction," *Quarterly Journal of Economics* 125:2 (2010), 549–589.
- Basu, Aparna, and Roland Wagner Döbler, "'Cognitive Mobility' or Migration of Authors between Fields Used in Mapping a Network of Mathematics," *Scientometrics* 91:2 (2012), 353–368.
- Borjas, George J., and Kirk B. Doran, "The Collapse of the Soviet Union and the Productivity of American Mathematicians," *Quarterly Journal of Economics* 127:3 (2012), 1143–1203.
- "Cognitive Mobility: Labor Market Responses to Supply Shocks in the Space of Ideas," *Journal of Labor Economics* 33 (2015), 5109–5145.
- Demoskop Weekly*. [demoscope.ru/weekly/ssp/sng\\_nac\\_89.php](http://demoscope.ru/weekly/ssp/sng_nac_89.php), accessed on February 27, 2013.
- Franzoni, Chiara, Giuseppe Scellato, and Paula Stephan, "The Mover's Advantage: Scientific Performance of Mobile Academics," NBER working paper (2012).
- Ganguli, Ina, "Immigration and Ideas: What Did Russian Scientists 'Bring' to the US?" Stockholm School of Economics working paper (2012).
- "Saving Soviet Science: The Impact of Grants When Government R&D Funding Disappears," *American Economic Journal: Applied Economics* (forthcoming).
- Graham, Loren, *Science in Russia and the Soviet Union: A Short History* (Cambridge: Cambridge University Press, 1993).
- Grogger, Jeffrey, and Gordon H. Hanson, "Attracting Talent: Location Choices of Foreign-Born PhDs in the US," NBER working paper (2013).
- Jewish Encyclopedia of Russia (Rossiyskaya Evreiskaya Entsiklopediya)*, (1995), trans. Josif and Vitaly Charny, [http://www.jewishgen.org/belarus/rje\\_a.htm](http://www.jewishgen.org/belarus/rje_a.htm).
- Jones, Charles I. "Growth and Ideas," in P. Aghion and S. Durlauf, eds., *Handbook of Economic Growth*, vol. 1, pt. 2 (New York: Elsevier, 2005).
- Jones, Charles I., and Paul M. Romer, "The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital," *American Economic Journal: Macroeconomics* 2:1 (2010), 224–245.
- Lucas, Robert E., "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22:1 (1998), 3–42.
- "Ideas and Growth," *Economica* 76:301 (2009), 1–19.
- Moser, Petra, Alessandra Voena, and Fabian Waldinger, "German Jewish Émigrés and U.S. Invention," *American Economic Review* 104 (2014), 3222–3255.
- Romer, Paul M., "Increasing Returns and Long-Run Growth," *Journal of Political Economy* 94:5 (1986), 1002–1037.
- "Endogenous Technological Change," *Journal of Political Economy* 98:5 (1990), S71–S102.
- Saul, Mark, "Kerosinka: An Episode in the History of Soviet Mathematics," *Notices of the AMS* 46:10 (1999), 1217–1220.
- Waldinger, Fabian, "Quality Matters: The Expulsion of Professors and the Consequences for Ph.D. Student Outcomes in Nazi Germany," *Journal of Political Economy* 118:4 (2010), 787–831.
- "Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany," *Review of Economic Studies* 79:2 (2012), 838–861.
- "Bombs, Brains and Science: The Role of Human and Physical Capital for the Creation of Scientific Knowledge," University of Warwick working paper (2013).