Immigration & Ideas: What Did Russian Scientists 'Bring' to the US?

Ina Ganguli*

October 18, 2012

Abstract

How have immigrant scientists contributed to the diffusion of knowledge in the United States? In this paper, I draw upon the end of the Soviet Union to study the link between immigration and the flow of scientific ideas. During Soviet times, the USSR was relatively "closed" to contact with researchers outside of the Eastern bloc. When the Soviet Union collapsed, there were new opportunities for Russian and U.S. scientists to communicate and for Russian scientists to emigrate. This "opening" and flow westward of Russian scientists and knowledge produced in the USSR allow me to examine the extent to which immigrants brought new ideas to the US. Following recent papers on localized knowledge diffusion that have examined individual inventor and scientist mobility within the US and citation patterns using rich panel datasets, I use a difference-in-difference approach to estimate the causal impact of migration on knowledge flows. I create a paper-pair-level dataset of papers published during the Soviet period by later migrants (treated papers) and non-migrants (controls) using matching techniques. I find that after a Russian scientist moved to the US, citations to his or her papers published during Soviet times increase relative to similar control papers authored by non-migrants. Differences by field suggest that migrants in the Life Sciences and Chemistry contributed more to the diffusion of ideas in the US than migrants in Physics and Mathematics.

*SITE, Stockholm School of Economics (e-mail: ina.ganguli@hhs.se). I appreciate data assistance from Raviv Murciano-Goroff. I gratefully acknowledge support from the SSRC Eurasia Program and the Harvard Labor & Worklife Program.

1 Introduction

Many scientists and engineers today leave their home countries, particularly from the developing world, to go abroad to pursue further training or to continue their careers (Weinberg, 2010; Hunter et al., 2009). Evidence shows that such immigrants are key contributors to U.S. innovation along a number of dimensions, usually measured by patents and publications (e.g. Hunt and Gauthier-Loiselle, 2010; Kerr and Lincoln, 2010). In addition to their direct contributions to research and innovation through their own papers and patents, the knowledge embodied in these workers has the potential to further benefit their new home country. Economic theory suggests that the human capital or ideas of these highly-skilled workers can "spill over" to others and increase their productivity; thus, having more highly-skilled workers in a country should increase the productivity of the existing ones.

How have immigrant scientists contributed to the diffusion of knowledge in the United States? In this paper, I draw upon the end of the Soviet Union to study the link between immigration and the flow of new ideas. During Soviet times, the USSR was relatively "closed" to contact with researchers outside of the Eastern bloc. When the Soviet Union collapsed, there were new opportunities for Russian and U.S. scientists to communicate and for Russian scientists to emigrate. This "opening" and flow westward of Russian scientists allow me to examine the extent to which immigrants brought new ideas to the US that served as inputs for knowledge production among US scientists.

Several recent studies have measured knowledge spillovers by estimating the impact of exogenous arrivals or departures of immigrant scientists on the publishing activity of their peers. Often limited to one or a few fields of science, the findings have been mixed. For example, Borjas and Doran (2012) use the influx of Soviet mathematicians to the US after the end of the USSR to examine the impact on the productivity of US mathematicians, finding negative productivity impacts on peers in subfields with greater overlap with Soviet mathematicians. Waldinger (2010) uses the Nazi expulsion of scientists during World War II and finds that the loss of high-quality scientists negatively impacted the productivity of younger cohorts of scientists, although it does not seem to affect their peers (Waldinger, 2012). Meanwhile, Moser et al. (2011) show that the arrival of German-Jewish emigre chemists positively impacted the patenting of US peers in their field. In a study focused on the diffusion of economic policy ideas across countries, Kogut and Macpherson (2011) show how the mobility of US-trained economists abroad impacted the adoption of economic policies in their destination country. In light of existing evidence and given the important role of human capital externalities as presumed inputs to innovation for economic growth (Romer, 1990), a greater understanding of the extent and channels through which immigrants contribute (or do no contribute) to the flow of new ideas into a country is needed.

In this paper I examine whether immigrants are a channel through which knowledge is diffused by building on an important literature examining the geographic localization of knowledge flows (Jaffe et al., 1993) and the role of immigrants in knowledge transfer (e.g. Kerr, 2008). The aim of my empirical strategy is to estimate the impact of Soviet scientist immigrants on U.S. scientists' 'ideas' by examining the citation trajectories of articles published during Soviet times (and before the scientist migrates) relative to control group of articles published by non-migrants.

A few recent papers have examined the impact of individual inventor and scientist mobility within the US on citation patterns using rich panel datasets of papers and patents (e.g. Azoulay et al., 2012; Singh and Agrawal, 2011). Following the basic methodology of these studies, I use a difference-in-difference approach and matching methods to create a product-level dataset of papers published during the Soviet period by subsequent migrants (treated papers) and non-migrants (control papers). Since the immigration decision and timing of migration is likely to be endogenous, the difference-in-differences strategy allows me to compare the forward citations of Soviet articles "treated" by migration of the author to the US to the citations of similar control papers not treated. The matching procedure matches on several characteristics constant over time, but also on citation trends before migration. Thus, I can compare US citations to very similar pre-1990 Soviet papers authored by migrants and non-migrants in both the pre- and post-move periods. If immigrants 'bring' their ideas with them, U.S. authors should be citing the migrants' papers more than the similar papers of Soviet scientists who did not move.

I use a unique panel dataset of Russian scientists, their publications and citing publications across many fields of science in my analysis. I identify and match Russian scientists to their publications, citations and affiliations during the pre- and post-Soviet periods using the Thomson Reuters ISI Web of Science database and use a unique name-matching technique to account for the transliteration of Russian names. I can observe the publishing activity of Russian scientists before and after the end of the USSR and I observe which scientists immigrated to the US.

I find that after a Russian scientist moved to the US, citations to his or her papers published during Soviet times increase relative to similar control papers authored by non-migrants. The increase is small, but significant and only occurs for a few years. However, given that the Soviet papers are relatively dated and that the estimated coefficients are at the paper-level, even this small increase suggests that immigrants do contribute to the flow of ideas across countries. In addition, some differences in the impact of migration across fields suggest that there may be some types of knowledge or conditions that differ by field that make ideas more likely to be diffused by migrants. In this case, it appears that Russian migrants in the Life Sciences and Chemistry tended to diffuse ideas in the US more than migrants in Physics and Mathematics.

The paper proceeds as follows. I provide some background about the historical context in Section 2. In Section 3 I describe the construction of the dataset. The empirical strategy and results follow in Section 4 and 5 respectively, and Section 6 concludes.

2 Background

While the USSR had a large scientific community, it was very "closed" to contact with researchers outside of the Eastern bloc. Scientists were rarely able to travel outside of the USSR, although there was substantial contact with scientists in other communist countries. In terms of access to journals and existing knowledge, scientists had access to the many Soviet journals as well as journals in the Eastern bloc, and were also able to order reprints from foreign journals from Moscow. However, there was usually a lag in receiving reprints and often the reprints were censored.

With the end of the USSR came many freedoms for scientists, including greater mobility and contact with the western world. Many scientists chose to move abroad to the United States, Israel or Europe to continue their careers. Others remained at home and sought opportunities to continue their research, in spite of the economic instability. Some, meanwhile, left science completely and pursued other career options. By 1993, Russian government estimates of the rate of change in researchers in the Academy of Science as a share of 1991 levels were -6.5%, -31% in industrial science and technology, and -35.2% in higher education institutions (Graham and Dezhina, 2008). In my analysis, I focus on scientists who would have primarily been in the Academy of Sciences or other institutions conducting basic research. The estimates of the number of scientists who left after the end of the USSR vary so widely that they are almost meaningless. Using data on Soviet scientists who had published in top journals during Soviet times, Ganguli (2012) shows that migrants tended to be younger, were more likely to have coauthored with a non-USSR scientist in the past, were more likely to come from Moscow, and were more likely to come from the upper deciles of the productivity distribution. There were also differences by scientific field, with Physics more represented among migrants than non-migrants, and Chemistry less represented.

One factor facilitating immigration and collaborations between Russian and Western researchers were the numerous foreign grant programs aimed at fostering international collaboration and exchange programs for Russian scientists. In the 1990s, the US Government spent $350 million per year on average to support science and technology cooperation with Russia, and $200 million per year specifically on joint research projects (Wagner et al., 2002). There was particular concern among western countries about the outflow of knowledge former Soviet scientists possessed about nuclear, chemical, and biological weapons. Worry that rogue nations or terrorist groups would recruit scientists to build weapons of mass destruction, or that they would sell their knowledge without leaving, was a key motivation for the creation of many western assistance programs.

During Soviet times, scientists in the US also had relatively little access to Soviet publications. However, the Institute for Scientific Information (ISI) had been indexing articles for the top Soviet journals for many years. However, the full text of these articles was not readily available to US scientists, especially translated into English. As Graham and Dezhina (2008) note, not many people outside the USSR knew Russian, and so US researchers were typically not aware of Soviet scientific achievement. Some of the scientific associations tried to facilitate members' knowledge about Soviet research, such as by publishing dictionaries of e.g. English-Russian mathematical terms, but these cases were rare. Scientists working at the time recalled being aware generally of the types of research being done, but that they typically did not know details about specific papers. After the collapse of the USSR, there was greater availability of Soviet publications in the US, but information took time to flow freely, and by the time publications were readily available on the web, many Soviet publications were already quite old. The primary channels through which Soviet papers became available to US researchers was most likely though contact with Russians at conferences and through Russian immigrants and international collaborations.

This isolationism is important for my empirical approach, as I limit the papers published by migrants and non-migrants to Soviet-era articles to exploit the fact that during Soviet times scientists in the US had relatively narrow access to Soviet publications. While there was greater availability of Soviet publications after 1991, it was still limited and Soviet-era research had already started to be outdated and was typically not translated into English. Moreover, after the end of the USSR, Russian scientists had more opportunities to publish in western journals, which would be readily accessible to scientists in the US. Since the articles of migrants and non-migrants published during Soviet times were thus subject to a similar level of isolationism, restricting the analysis to subsequent citations to these papers lessens the worry that other factors correlated with migration are driving the results rather than migration itself[1]

3 Data

To proceed with the analysis I constructed a dataset that linked the Soviet-era publications of Russian scientists to the timing and location of their post-Soviet migration behavior, as well as to the papers published in the US (not by them) that subsequently cited their Soviet-era publications.

I first identified a sample of scientists who were "doing science" in Russia around the time of the Soviet collapse. I use publication data from the Thomson Reuters ISI Web of Science (ISI)[2] to create a sample of Russian scientists who were actively doing scientific research near the end of the Soviet Union in 1991. I identified the top Soviet

[1]While there was a shift towards greater openness after 1985 during perestroika, interviews with scientists suggest there were no dramatic changes in the availability of Soviet/US publications.

[2]Web of Science ® prepared by THOMSON REUTERS ®, Inc. (Thomson®), Philadelphia, Pennsylvania, USA: © Copyright THOMSON REUTERS ® 2010. All rights reserved.

and Russian language journals in the ISI database and extracted all authors publishing in these journals between 1986-1994 (see Appendix Table B2 for the list of Russian/Soviet journals used to create the sample). The ISI database includes over 100 top journals of the former USSR and Russian language journals. It includes journal backfiles to 1900, however journals entered the database at different times. By the 1970s, most of the Russian journals in the database in later years had entered the database.

I next identified the subset of authors who had an address that included a city in the former Russian Republic of the Soviet Union. I dropped any individuals with a foreign address before 1990. I further restricted to authors who "stayed in science" after 1991 and I could identify their location, meaning they published at least one article after the end of the USSR through 2008. For each scientist I have a record of their publications from the year they first enter the ISI database through 2008. For each paper, I know all the basic information, such as the year it was published, the journal, author and corresponding author addresses, the number of coauthors, each author's position, subject categories of the journal, and number of pages. Appendix A provides a full description of the preparation of the publication data, including information on transliteration and name matching and assigning scientific fields to individuals.

3.1 Migrants

To determine who migrated to the US, where and in which year, I use information from the author addresses on each publication. From the addresses I can identify when someone first published in the US and in which city. If I observe a Russian scientist in my sample who publishing in a US city after 1991, I define them as a "migrant" and I define the year of migration as the year they first publish in the US. While many Russian scientists may have moved later within the US (or back to Russia), I restrict the analysis to the first time I observe a scientist arriving in the US.

Given that the date of the first publication in the US is a very noisy measure of the year of migration, a more accurate assessment of the migration year would be preferred using information from CVs or other data not culled from publications. Unfortunately, the lack of CVs and other information for the scientists in my sample (via websites, etc.) makes it difficult to determine a more exact move year.

The full sample of migrants I observe moving to the US between 1992-2002 is 809. For the paper-level analysis, I restrict this sample to migrants who are ever cited by a US paper, which reduces the number of migrants to 535. Figure 1 shows the total number of migrants I observe by year. It shows that there is variation across time, but that there were inflows of migrants throughout the 10-year period. The peak is in 1995 and then again an increase in 1999 (probably resulting from the 1998 Russian economic crisis). Figure 2 shows the distribution of the migrants by the main scientific fields. Physics (39%) and Life Sciences (30%) make up the largest share of migrants in the sample.

3.2 US Citations to Soviet-era Publications

I next matched all articles published between 1980-1990 by the Russian scientists in the sample to the papers that subsequently cited them in the US. After making this link, I could calculate the number of US citations to Soviet papers by year, US city, and field of the Russian scientist. Note that I exclude self-cites and citing papers that include a Russian coauthor (to prevent including cites from papers by the immigrants themselves or with Russian collaborators).

I limit the Soviet articles to those published from 1980-1990 to exploit the fact that during Soviet times scientists in the US had little access to Soviet publications. As discussed in the previous section, while there was greater availability of Soviet publications after 1991, it was still limited and Soviet-era research had already started to be outdated, if only due to the lengthening lag compared to the speed as which science advances. Since the articles of migrants and non-migrants published during this time were thus all subject to the same level of isolationism, restricting the analysis to subsequent citations to these papers lessens the worry that other factors correlated with migration are driving the results rather than the mobility itself.[3] Moreover, there was a large economic collapse that accompanied the end of the USSR, and many scientists exited the science sector (Ganguli, 2011). Restricting the publications to the pre-transition period further prevents confounding the impact of migration from other factors like the economic conditions that may have differentially impacted on the quality of migrants' vs. non-migrants' research.

However, a trade-off to using articles produced only from 1980-1990 is that the number of citations to these articles originating in the US is low, even in the post-Soviet period, partly due to to the growing lag from publication mentioned before. Thus, the variation in the number of citations originating in the US is much lower than if I had included citations to more contemporary articles of migrants and non-migrants in the analysis.

[3]While there was a shift towards greater openness after 1985 during perestroika, interviews with scientists suggest there were no dramatic changes in the availability of Soviet/US publications.

In Figure 3, I show the distribution of the year of publication of all Soviet-era papers in my sample. There is a slight increase in articles published from 1988-1990, but the articles are otherwise evenly distributed across the years. Figure 4 presents the total number of total US citations to Soviet articles published 1980-1990 from 1992-2002. The number of citations is declining until 1997, then peaks, and then begins to decrease. In Figure 5 I show the total citations from 1992-2002 by the originating city (for the top cities only). The full sample has 292 total cities. The top cities are not a surprise, mainly large cities with large universities. However, there is a great deal of variation across the cities and also reflect areas that specialize in specific fields (e.g Los Alamos).

Descriptive statistics for the full sample of Soviet-era articles (1980-1990) published by migrants and non-migrants are presented in Table 1. There are 19,752 articles published by non-migrants compared to 2,570 articles by migrants. Covariates at both the article-level and scientist-level are included. Clearly, migrants look very different from non-migrants. Their articles are slightly more recent on average and have fewer coauthors. Migrants are also more productive during the Soviet period. There are many differences across the fields, with a migrants more likely to be in the Life Sciences, Mathematics and Physics than non-migrants. These differences across fields also likely contribute to the the productivity differences. Finally, migrants are more likely to come from Moscow than non-migrants.

4 Empirical Approaches

4.1 City-Field-Level Panel

I begin with an analysis of the contribution of immigrants to the flow of ideas using a panel of city-field level data on the number of migrants arriving to US cities by field after the end of the USSR and the growth in the number of citations originating in US cities by field in that period. This approach exploits variation across cities, fields and time in the number of immigrant scientists arriving to the US from Russia to identify the impact of immigration on the flow of ideas. Following a basic growth model specification, I estimate the following:

$$\Delta y_{ijt} = \gamma y_{ijt-1} + \beta_1 Migrants_{ijt} + \theta_i + \rho_j + \lambda_t + \psi_{jt} + \varepsilon_{ijt} \quad (1)$$

where Δy_{ijt} is the growth rate in US citations to Soviet papers published by authors in US city i in field j in year t and y_{ijt-1} is the number of citations in the previous year. *Migrants* is the number (inflow) of migrants, θ_i are city fixed effects, ρ_j are field fixed effects, λ_t are year fixed effects and ψ_{jt} represents field x year fixed effects. The field and year interactions control for changes in the fields over time, and robust standard errors are clustered at the city-field level. The coefficient on *Migrants*, β_1, provides an estimate of the growth in citations to Soviet-era papers in a given field in a particular US city due to the presence of additional Russian migrants.

There are several concerns with this approach. First, the citation at the city-field-year level contains many zero values since no one in a given year would publish a paper citing a Soviet article from a particular field. This is not surprising given that papers from the Soviet period would be cited less in the US in general due to the lack of communication during the Soviet period, and then further disaggregating them by city and field leads to many empty cells.

More importantly, despite the variation in the number of immigrants arriving in the US over this period and across cities and fields, there are concerns about endogeneity. The coefficient on *Migrants* would be biased upwards if immigrants are more likely to move to areas where Soviet papers would be cited more anyway in that year or if there are unobserved differences that change over time that may be influencing the citation behavior of US scientists in particular cities and/or fields.

An advantage of this approach is that the outcome measure of the growth of citations includes all Soviet papers in my sample (described in the Data section), rather than only the papers published by the immigrant themselves. Typically, studies that estimate the impact of individual mobility on the diffusion of ideas limit the analysis to papers or patents published by the focal (moving) individual. While my main analysis also uses this approach (described next), it is also useful to see how the greater stock of knowledge embodied in an individual may be diffused after a move. Since Soviet scientists presumably had knowledge about not only his/her own research, but also those of their colleagues from the USSR, this approach can more fully capture the potential spillovers.

Moreover, this approach does not suffer as much from the concern noted in Jaffe et al. (1993) about whether paper citations represent true knowledge spillovers if scientists e.g. cite a friend or colleague "just to be nice" because it is very low cost to do so, and therefore is probably a more accurate measure of knowledge flows in this sense.

Note that this measure also excludes self-cites and US citing papers including a Russian collaborator. Finally, in this analysis, the number of immigrants arriving in the US is not limited to only those scientists who were cited at some point by US authors (during the pre- or post-Soviet period), which is the case for the following paper-pair level analysis.

4.2 Difference-in-Difference Analysis

Next, I use a difference-in-difference analysis at the matched paper-pair level to provide causal evidence on the contribution of immigrants to the flow of ideas at a very micro-level. I match each paper by a scientist who migrated to a scientist who did not emigrate based on their observable characteristics and the characteristics of the paper. To ensure that the papers treated by migration are comparable to control papers, I use a nonparametric matching method, "Coarsened Exact Matching" (CEM) (Blackwell et al., 2009), which has been used in the recent economics of science and innovation literature to create matched control groups for scientists, articles, and patents.

The basic approach of CEM is to temporarily coarsen each covariate used in the matching process, create unique strata based on the coarsened values of the covariates, assign each observation to a stratum, and then drop any observations in strata in which there isn't a control observation for each treatment observation (Blackwell et al., 2009). This implies that any treated observations for which there is no control observation in the strata are dropped. At the end of the process, there are an equal number of treatment and control observations that are balanced on the covariates selected to match on.

Using CEM, I create pairs of matched papers published by migrants and non-migrants who look very similar on observables at the time the migrant moves. I do this by matching on both scientist-level covariates and paper-level covariates to ensure that migrants and non-migrants look very similar in terms of productivity, renown and openness as well as to ensure that the papers of migrants and non-migrants matched look very similar, including on their pre-migration citation trends.

The scientist-level covariates used in the matching process include scientific field, total citations to articles pre-1990 (not restricted to citations originating in the US), career age (coarsened to several categories), having had a foreign coauthor pre-1990, and originally from Moscow. I also match on article covariates, including the year of publication, the number of coauthors and the pre-move citations originating in the US.[4]

Following Azoulay et al. (2012) I use the following specification to test for the impact of migration on citations at the matched paper-pair level[5]:

$$Cites_{mt} - Cites_{nt} = \beta_0 + \beta_1 PostMigration_{kt} + f(Age_{kt}) + \phi_{mn} \tag{2}$$

The dependent variable is the difference in total citations in year t between all US citations received to a paper "treated" by migration and the matched control paper, where m denotes an article written by a migrant and n is the matched paper written by a non-migrant. $PostMigration$ is an indicator variable for the years after the migrant (focal) scientist k moves to the US, $f(AGE)$ is a function of scientist k's age, and ϕ_{mn} are the paper-pair fixed effects. Thus, β_1 reflects changes in the pair's citation rate following the

[4]I also tried to refine the match by including additional paper-level covariates such as the the journal and the author's position, but including these in the matching process greatly increased the number of strata and decreased the match rate significantly.

[5]Singh and Agrawal (2011) use a different specification for the difference-in-difference estimation. They use patent-level data for focal and control inventors and include patent fixed effects. I also present results in the appendix using this alternate specification.

migration of scientist k. Robust standard errors are clustered at the scientist-level. All results I present use ordinary least squares (OLS) estimation.

Note that the dependent variable does not disaggregate the US citations by city, as there are too many zero values if I only measure citations originating in a particular city. Thus, these results consider the geographic localization of knowledge at a larger scale (country) than often studied in the literature.

5 Results

5.1 City-Field Panel Results

I first present the results of the regressions relating the growth in citations to Soviet articles by city and field to the number of migrants in a given year. These regressions should be interpreted with caution due to endogeneity concerns, as unobserved factors maybe be correlated with the migrants arriving in cities and fields and the citing behavior of the US scientists in those places. However, they provide some broader suggestive evidence on the link between migration and knowledge flows before moving on to providing causal evidence at a much more micro-level approach with the difference-in-differences analysis.

The coefficients in Table 3 provide the coefficients on the number of migrants from equation (1) using the city-field panel data for years 1992-2002. Column (1) shows that there is a positive and significant relationship between the number of new migrants in a given city, field and year and the growth in citations to Soviet-era articles. The coefficient suggests that each additional migrant is associated with approximately a 20 percentage point increase in the citation growth rate. When New York and Cambridge, which had the largest inflows of Soviet scientists, are excluded from the regression in Column (2), the coefficient becomes smaller, but is still statistically significant. Since access to Soviet-era research through other means was likely to have happened in the early part of the post-Soviet period, I also exclude 1992-1995 to see if the relationship still holds. Column (3) shows that the coefficient becomes larger when restricting the panel to later years. Finally, I interact the number of migrants with the scientific field dummies to see if there are differences by field. The coefficients show that the effect is primarily driven by growth in citations in Life Science and Chemistry. The effect is smaller for Physics and even negative for Mathematics.

5.2 Matched Paper Pair Difference-in-Difference Results

Table 2 shows the descriptive statistics for the matched sample of 2,448 total papers (or 1,224 pairs). This implies a match rate of 47.6% of the migrants' articles.[6] Some of the covariates presented were used in the matching process and others were not. Clearly, the match appears to have worked quite well in pairing similar articles published by similar authors. There are no significant differences for any variables across the migrant and non-migrant groups. Note that there are 422 migrants in the sample but 864 non-migrants, since individual papers by a migrant can be matched to control papers written by non-migrants.

The main results are visible in Figure 6. This graph shows the average difference in US citations between the paired papers (solid line) with a 95% confidence interval (dashed line) by the time to migration (in years). The trend shows that the match was

[6] Azoulay et al. (2012) report a match rate of 25.61%, although they begin with more mover articles and match on more article-level covariates, making their matching process more stringent.

effective in pairing articles with similar pre-migration citation trends. A positive effect of migration on citations in the US can be seen beginning after the second year post-migration. The increase is significant for a few years before fading out. The timing of the increase suggests that there is a lag between when a migrant moves to the US and when the Soviet-era paper starts to get cited. Given that I am measuring citations to the Soviet publication originating anywhere in the US, this would be consistent with models of localized knowledge diffusion, since it would take time for the knowledge embodied in the migrant to diffuse throughout the country.

The regression results corresponding to equation (2) are presented in Table 4.[7] These regressions include paper-pair fixed effects and are OLS regressions with robust standard errors clustered at the scientist-level. Column (1) shows that the coefficient on *Post-Migration* is positive and significant, although the magnitude is quite small. However, considering that these papers are already quite dated, it is still notable that there is a significant effect and suggests that migrants did contribute to the diffusion of ideas in the US. Column (3) widens the window around the year of migration to 8 years before and after. The coefficient is slightly smaller and still significant.

In Columns (2) and (4), I include interactions of the *Post-Migration* indicator with the scientific field. In the 5-year window, the effect for Mathematics is smaller compared to the other fields, and even slightly negative, suggesting that control articles were cited more. While the negative effect is puzzling, it could reflect the differences Borjas and Doran (2012) point to regarding the differential distribution of US vs. Soviet

[7]The results using Singh and Agrawal (2011)'s specification on paper-level data and paper fixed effects are in Appendix Table B1. The basic result of an increase in citations to the migrant's papers post-migration holds, but it is only significant at the 10% level.

mathematicians across subfields. A possibile story consistent with these results would be that the migrants passed on the ideas of their Russian colleagues in different subfields who stayed in Russia. In the 8-year window, however, the negative effect for Mathematics disappears, and the impact of migration for physicists decreases relative to other fields, so that the effect is close to 0. Given that physics research institutes were among the first to get Internet access across Russia, this might have led to greater flows of knowledge to the US in the field and could have made the role of migration in the diffusion of ideas less important. Interestingly, these results are consistent with the city-field-level panel results in the previous section, which also showed a negative effect for Mathematics and a smaller effect for Physics.

6 Summary & Conclusions

In this paper I examine whether immigrants are a channel through which scientific knowledge is diffused. I use the end of the Soviet Union to study the link between immigration and the flow of new ideas by exploiting the fact that during Soviet times, the USSR was relatively "closed" to contact with foreign researchers and to emigration until it collapsed in 1991. I use a unique panel dataset of Russian scientists, their publications and the US publications that cite them across many fields of science to create a matched paper-pair-level dataset of migrants' Soviet-era papers (treated) and non-migrants' papers (control papers).

My results show that after a Russian scientist moved to the US, citations to his or her papers published during Soviet times increase relative to similar control papers authored by non-migrants. The increase is small, but given that the Soviet papers would have been relatively dated and that the estimated coefficients are at the paper-level, even this small increase suggests that immigrants do contribute to the flow of ideas across countries.

In addition, some differences in the impact of migration across fields suggest that there may be certain types of knowledge or some conditions that differ by field that make ideas more likely to be diffused by migrants. While a greater understanding is needed of the differences between the Soviet scientific fields that may have led to a greater impact of migration on the flow of knowledge across fields, these results suggest that there are promising avenues for further research on identifying conditions more conducive to knowledge flows through migration than others.

The evidence I provide in this paper contributes to the literature on the impact of geography on the flow of knowledge with new evidence on the contribution of high-skilled immigrants to these flows. Although there are issues to consider when making the link beteween paper-to-paper citations and knowledge flows, this 'paper trail' provides useful clues as to the ways that immigrants can contribute to bringing new ideas into a country.

While in this paper I do not not address whether the new ideas brought to the US by Russian scientists impacted the productivity of native scientists, the findings do point to the possible channels through which such productivity increases could occur and in which fields of science. Moreover, while decreased travel and communication costs seem to be blurring borders and decreasing the importance of geography for transfers of knowledge, these findings show that during a time of sudden decreases in communication costs, geographic proximity through migration can still be an important channel for knowledge flows.

References

Azoulay, P., Zivin, J., and Sampat, B. (2012). The Diffusion of Scientific Knowledge across Time and Space. *The Rate and Direction of Inventive Activity Revisited*, page 107.

Blackwell, M., Iacus, S., King, G., and Porro, G. (2009). CEM: Coarsened Exact Matching for Stata. *Stata Journal*, 9(4):524–546.

Borjas, G. and Doran, K. (2012). The Collapse of the Soviet Union and the Productivity of American Mathematicians. *The Quarterly Journal of Economics*, 127(3):1143–1203.

Ganguli, I. (2011). Saving Soviet Science: The Impact of Grants When Government R&D Funding Disappears. *Harvard University, unpublished manuscript.*

Ganguli, I. (2012). Location and Scientific Productivity: Evidence from the Soviet "Brain Drain". *Harvard University, unpublished manuscript.*

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

Hunt, J. and Gauthier-Loiselle, M. (2010). How much does immigration boost innovation? *American Economic Journal: Macroeconomics*, 2(2).

Hunter, R., Oswald, A., and Charlton, B. (2009). The Elite Brain Drain. *The Economic Journal*, 119(538):F231–F251.

Jaffe, A., Trajtenberg, M., and Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, 108(3):577.

Kerr, W. (2008). Ethnic scientific communities and international technology diffusion. *The Review of Economics and Statistics*, 90(3):518–537.

Kerr, W. and Lincoln, W. (2010). The supply side of innovation: H-1b visa reforms and us ethnic invention. *Journal of Labor Economics*, 28(3):473–508.

Kogut, B. and Macpherson, J. (2011). The mobility of economists and the diffusion of policy ideas: The influence of economics on national policies. *Research Policy*, 40(10):1307–1320.

Leydesdorff, L. and Rafols, I. (2009). A Global Map of Science Based on the ISI Subject Categories. *Journal of the American Society for Information Science and Technology*, 60(2):348–362.

Moser, P., Voena, A., and Waldinger, F. (2011). German-jewish emigres and us invention. *SSRN Working Paper 1910247*.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, 98(S5):71.

Singh, J. and Agrawal, A. (2011). Recruiting for ideas: How firms exploit the prior inventions of new hires. *Management Science*, 57(1):129–150.

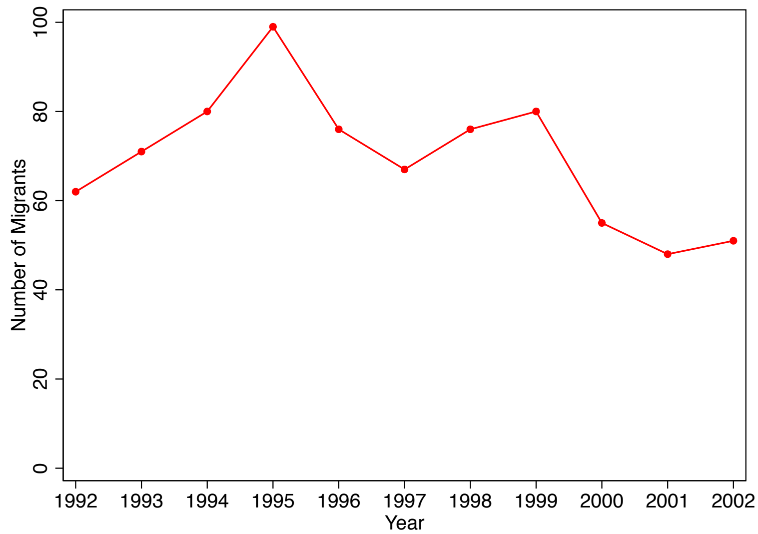
Wagner, C., Brahmakulam, I., Peterson, D., Staheli, L., and Wong, A. (2002). *US Government Funding for Science and Technology Cooperation with Russia*. Rand Corporation, Santa Monica.

Waldinger, F. (2010). Quality Matters: The Expulsion of Professors and the Consequences for PhD Student Outcomes in Nazi Germany. *Journal of Political Economy*, 118(4):787–831.

Waldinger, F. (2012). Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany. *The Review of Economic Studies*, 79(2):838–861.

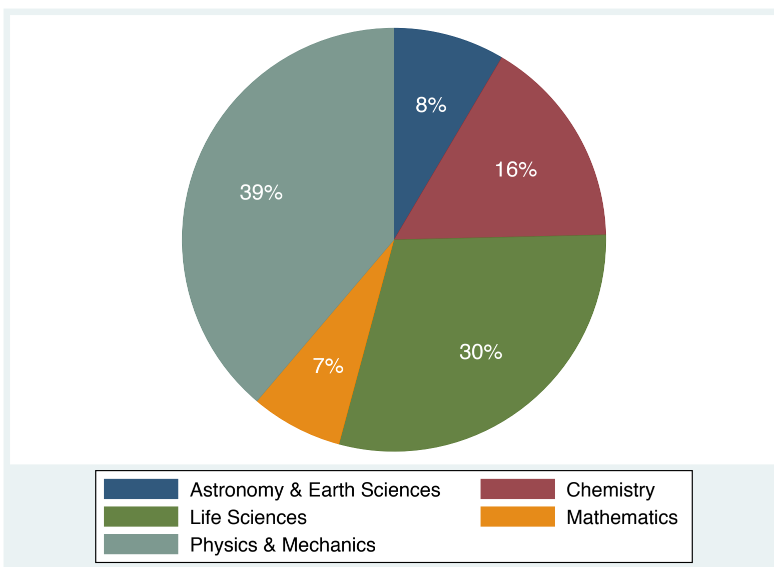
Weinberg, B. (2010). Developing Science: Scientific Performance and Brain Drains in the Developing World. *Journal of Development Economics*.

Figure 1: Flows of Russian Migrant Scientists, post-1992



Notes: Only includes Russian scientists who had published in a top Soviet journal from 1986-1990 in the ISI database.

Figure 2: Migrants by Field, post-1992



Notes: Only includes Russian scientists who had published in a top Soviet journal from 1986-1990 in the ISI database.

Figure 3: Year of Publication of Soviet-Era Papers (1980-1990)

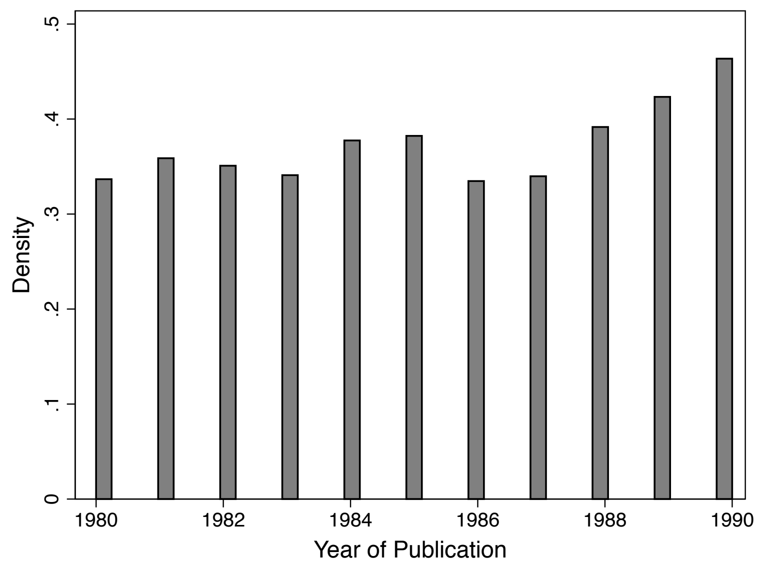
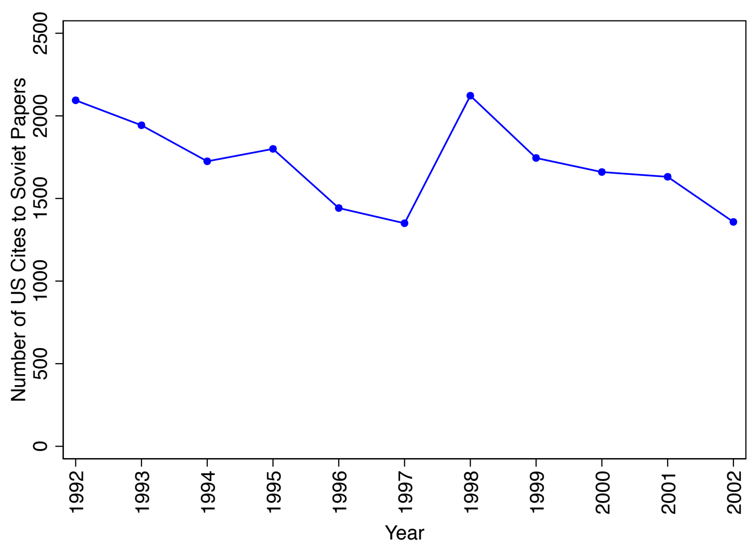
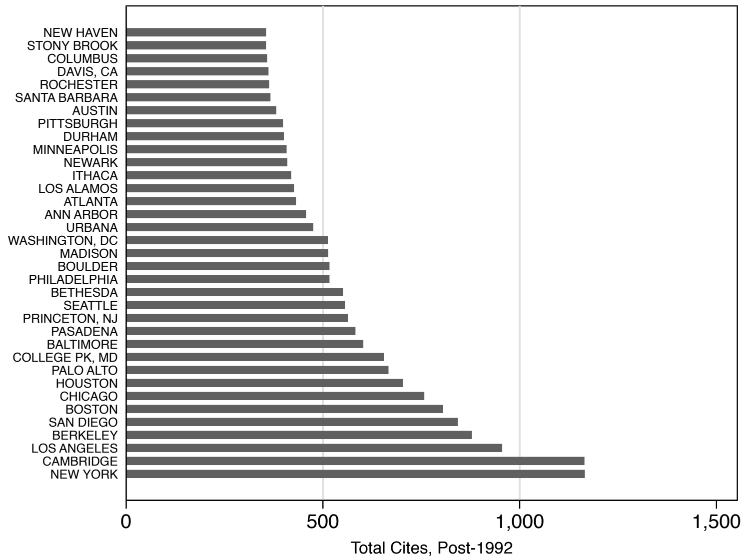


Figure 4: US Citations to Soviet Papers (1980-1990), post-1992



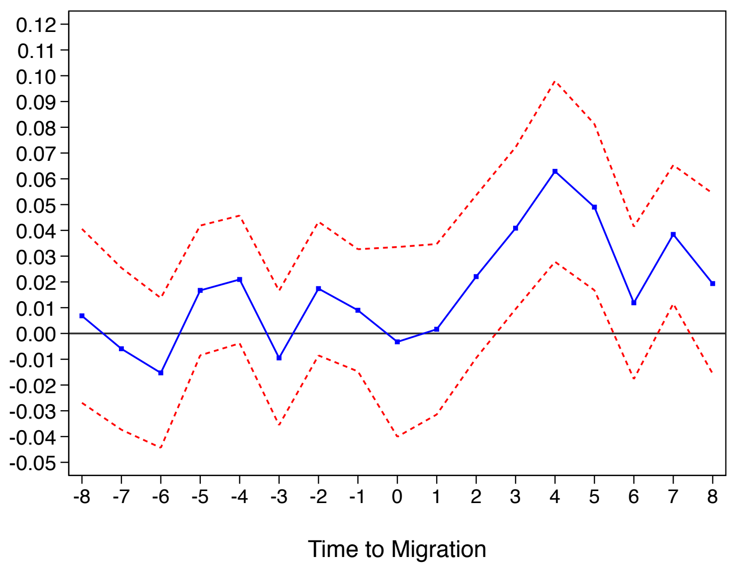
Notes: Citations are to papers by Soviet authors published 1980-1990.

Figure 5: Cities with Most Citations to Soviet Papers, post-1992



Notes: Citations are to papers by Soviet authors published 1980-1990.

Figure 6: Effect of Migration on Article Citation Rates



Notes: Difference in US citations between pairs of migrants and control matched articles written between 1980-1990.

Table 1: Summary Statistics, Full Sample of Articles

	Non-Migrants	Migrants	Difference
Year Article Published	1985.19	1985.77	-0.586***
No. Coauthors	4.40	3.77	0.630***
No. Pubs 1980-1990	34.66	36.07	-1.405**
No. Cites 1980-1990	327.03	783.19	-456.154***
Year of First Pub.	1968.12	1970.67	-2.552***
Non-USSR Coau. Pre-90	0.42	0.59	-0.172***
Astronomy & Earth Sci.	0.08	0.05	0.021***
Chemistry	0.29	0.15	0.141***
Life Sciences	0.26	0.36	-0.093***
Mathematics	0.02	0.05	-0.025***
Physics & Mechanics	0.35	0.39	-0.044***
Moscow Origin	0.69	0.78	-0.087***
Observations (Articles)	19752	2570	

Notes: Stars indicate the results of t-tests for the equality of means.

Table 2: Summary Statistics for CEM Matched Paper Pairs

	Control	Treatment	Difference
Article Age	10.82	10.82	0.000
No. Coauthors	3.86	3.76	0.104
No. Pubs 1980-1990	36.82	35.32	1.496
No. Cites 1980-1990	409.43	419.36	-9.934
Year of First Pub.	1970.46	1970.72	-0.256
Non-USSR Coau. Pre-90	0.50	0.50	0.000
Astronomy & Earth Sci.	0.05	0.05	0.000
Chemistry	0.18	0.18	0.000
Life Sciences	0.31	0.31	0.000
Mathematics	0.03	0.03	0.000
Physics & Mechanics	0.44	0.44	0.000
Moscow Origin	0.79	0.79	0.000
Article's Baseline Stock of US Citations	1.32	1.33	-0.016
Observations (Articles)	1224	1224	

Notes: Stars in the "Difference" column would indicate the results of t-tests for the equality of means.

Table 3: City-Field-Year Regressions: Migrant Inflows and Growth of Citations

	(1)	(2)	(3)	(4)
No. Migrants	0.197**	0.211**	0.239**	0.304**
	(0.048)	(0.050)	(0.074)	(0.058)
Soviet Cites in t-1	-0.024**	-0.027**	-0.029**	-0.025**
	(0.002)	(0.002)	(0.003)	(0.002)
No. Migrants X Astronomy & Earth Sci.				-0.381**
				(0.118)
No. Migrants X Chemistry				0.123
				(0.228)
No. Migrants X Math				-0.602**
				(0.101)
No. Migrants X Physics & Mechanics				-0.133+
				(0.077)
Constant	-0.192*	-0.232**	-0.302**	-0.321**
	(0.082)	(0.084)	(0.085)	(0.121)
City FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Field FE	Yes	Yes	Yes	No
Year X Field FE	Yes	Yes	Yes	Yes
R2	0.056	0.058	0.072	0.058
Nb. of Obs.	16,060	15,950	10,220	16,060
Nb. Clust.	1,460	1,450	1,460	1,460

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Notes: Dependent variable is the growth rate in citations. Estimates are ordinary least squares (OLS). Observations are at the city-field-year-level for years 1992-2002. Column (2) excludes New York and Cambridge; (3) only includes years 1996-2002. Robust standard errors are clustered at the city-field level (292 cities and 5 fields).

Table 4: CEM Matched Article Pair Regressions

	5-yr (1)	5-yr (2)	8-yr (3)	8-yr (4)
Post-Migration	0.030*	0.050*	0.028*	0.055*
	(0.013)	(0.025)	(0.013)	(0.023)
Field (Life Science omitted)				
Post-Migration X Astronomy & Earth Sci.		0.024		-0.004
		(0.041)		(0.033)
Post-Migration X Chemistry		-0.024		-0.036
		(0.031)		(0.027)
Post-Migration X Math		-0.110*		-0.074
		(0.052)		(0.052)
Post-Migration X Physics & Mechanics		-0.034		-0.042+
		(0.028)		(0.024)
Constant	0.013	0.013	-0.000	0.001
	(0.037)	(0.036)	(0.026)	(0.026)
Pair FE	Yes	Yes	Yes	Yes
Age Cat Dummies	Yes	Yes	Yes	Yes
R2	0.001	0.002	0.001	0.002
Nb. of Obs.	13,016	13,016	18,898	18,898
Nb. Clust.	422	422	422	422

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Notes: Dependent variable is the difference in citations originating in the US for treatment and control matched pairs. Estimates are ordinary least squares (OLS). Robust standard errors are clustered at the scientist-level. Columns (1) and (2) are for 5-year windows around the year of migration (between 1992-2002) and (3) and (4) are 8-year windows.

Appendix A Data Description

In this appendix I include further details about the construction of the dataset.

A.1 Publication Data

To create the sample of scientists who published in the top Soviet and Russian journals, I first identified these journals in the Thomson Reuters ISI Web of Science (ISI). I relied primarily on the list of journals identified by the International Science Foundation (ISF) as "qualifying" journals for potential grantees in Appendix Table B2 (see Ganguli (2011) for more information about the ISF.) In addition to searching journal titles for these titles, I identified the English language translations of these titles. I also checked the journals identified as Russian language journals using the language field in the ISI in case I missed articles published using different English translations.

The ISI publications are not associated with one scientific field, but with rather many scientific subject categories. In order to assign a scientific field to each individual scientist, I did the following. First, I assigned a likely broader scientific field to each of the 221 unique subject categories in the sample of publications based on the search of scientist names. The fields I assigned were one of the 7 major scientific field identified by the ISF (Astronomy, Chemistry, Earth Sciences, Life Sciences, Mathematics, Mechanics, and Physics). Note that when the grantees applied for the grants, they were asked to choose from one of these scientific fields and a number of subfields. Many of the subject categories clearly belonged to a scientific field, e.g. "Cell Biology" was coded as Biology and "Chemistry, Organic" was coded as "Chemistry". For other fields, I used resources that listed field codes along with the scientific field associated with it. For example, Thomson Reuters Journal Citation Report and Essential Science Indicators list subject categories with the broader scientific field, e.g. "Acoustics" is listed under "Physics".

I also compared field codes with the results of analysis presented in Leydesdorff and Rafols (2009), who use exploratory factor analysis of the matrix of field codes in the ISI database to determine the disciplines associated with each subject category. If a subject category could belong to more than one scientific field, I did not code it. Then, to assign a scientific field to each publication, I chose the most common scientific field among the subject categories. Then, for each scientist, I chose the most common scientific field among all the publications he/she published.

A.2 Transliteration & Name Matching Issues

A challenge in matching the scientists to publication data is how names from the Cyrillic alphabet are transliterated into the Latin alphabet. Using a name dictionary Polyglossum 3.71 created by ETS Publishing House (Moscow) that is based on several official standards for transliterations (e.g. ISO 9-1995, OVIR of Russia regulations), I identified possible spellings for each last name and searched for each variant in the publication databases. For example, an example of a surname in my sample in the Latin alphabet is Kuznetsov. This Cyrillic name (Кузнецов) has multiple transliterations, which I identified with the name dictionary:

Кузнецов:
Kuznetsov
Kuznecov
Kouznetsov (*OVIR USSR)

where "OVIR USSR" is the transliteration standard used by the "Office of Visas and Registration" of the USSR. Note that this is not an issue for many names, such as Ivanov, or names from the Baltic countries, as in these countries the languages use the Latin alphabet.

Additionally, typical name ambiguity issues arise, including common names (such as Ivanov, like Smith in the U.S.). I exclude these names from the analysis. In the original grantee list, approximately 4% of the grantees had non-unique names. I also trim the data by excluding name with more than 500 publications during the period, since the likelihood of common names is higher for these occurrences. The results do not change significantly when these observations are included.

Appendix B Additional Figures & Tables

This section includes additional figures and tables mentioned in the text.

Table B1: CEM Matched Regressions, Alternate Specification at Paper-level

	(1)	(2)
Post-Migration	0.019^{+}	0.019^{+}
	(0.010)	(0.010)
Constant	0.190^{**}	0.183^{**}
	(0.059)	(0.060)
Pair FE	Yes	Yes
Year FE	Yes	Yes
Cite lag FE Dummies	Yes	Yes
Age Cat Dummies	No	Yes
R2	0.005	0.005
Nb. of Obs.	37,796	37,796
Nb. Clust.	1,283	1,283

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Notes: This difference-in-difference specification follows Singh and Agrawal (2011) and the data is at the scientist-paper-level. The dependent variable is number of citations in a given year and the model includes paper fixed effects. Estimates are ordinary least squares (OLS). Robust standard errors are clustered at the scientist-level. There are 1,283 scientists in the sample, of which 424 are migrants.

Table B2: List of Qualifying Journals from the former USSR

1. AKUSTICHESKII ZHURNAL
2. ANTIBIOTKI I KHIMIOTERAPIYA
3. ARKHIV PATOLOGII
4. ASTRONOMICHESKII ZHURNAL
5. BIOLOGICHESKIE MEMBRANY
6. BIOLOGIYA MORYA
7. BIOORGANICHESKAYA KHIMIYA
8. BIOFIZIKA
9. BIOKHIMIYA
10. BYULLETEN EKSPERIMENTALNOI BIOLOGII I MEDITSINI
11. VESTNIK AKADEMII MEDITSINSKIKH NAUK SSSR
12. VESTNIK AKADEMII NAUK SSSR
13. VESTNIK MOSKOVSKOGO UNIVERSITEfA SERIYA F1Z1KA I ASTRONOMIYA
14. VESTNIK MOSKOVSKOGO UNIVERSITETA SERIYA KHIMIYA
15. VESTNIK MOSKOVSKOGO UNIVERSITEfA SERIYA MATEMATIKII MEKHAN1KI
16. VOPROSY VIRUSOLOGII
17. VOPROSY MEDITSINSKOI KHIMII
18. VOPROSY ONKOLOGII
19. VYSOKOMOLEKULYARNYE SOEDINENIYA SERIYA A
20. VYSOKOMOLEKULYARNYE SOEDINENIYA SERIYA B
21. GENETIKA
22. GEOMAGNETIZM I AERONOMIYA
23. GEOTEKTONIKA
24. GEOKHIMIYA
25. DIFFERENTSIALNYE URAVNENIYA
26. DOKLADY AKADEMII NAUK BSSR
27. DOKLADY AKADEMII NAUK SSSR
28. ZHURNAL ANALITICHESKOI KHIMII SSSR
29. ZHURNAL VYSSHEI NERVNOI DEYATELNOSTI IMENI I P PAVLOVA
30. ZHURNAL MIKROBIOLOGII EPIDEMILOGII I IMMUNOLOGII
31. ZHURNAL NEORGAATJCHESKOI KHIMII
32. ZHURNAL OBSHCHEI BIOLOGII
33. ZHURNAL OBSHCHEI KHIMII
34. ZHURNAL ORGANICHESKOI KHIMII
35. ZHURNAL STRUKTURNOI KHIMII
36. ZHURNAL TEKHNICHESKOI FIZIKI
37. ZHURNAL FIZICHESKOI KHIMII
38. ZHURNAL EVOLYUTSIONNOI BIOKHIMII I FIZIOLOGII
39. ZHURNAL EKSPERIMENTALNOI I TEORETICHESKOI FIZIKI
40. ZOOLOGICHESKY ZHURNAL
41. IZVESTIYA AKADEMII NAUK SSSR SERIYA BIOLOGICHESKAYA
42. IZVESTIYA AKADEMII NAUK SSSR SERIYA GEOLOGICHESKAYA
43. IZVESTIYA AKADEMII NAUK SSSR SERIYA FIZIKI ATMOSFERI I OKEANA
44. IZVESTIYA AKADEMII NAUK SSSR SERIYA FIZIKI ZEMLI
45. IZVESTIYA AKADEMII NAUK SSSR SERIYA FIZICHESKAYA
46. IZVESTIYA AKADEMII NAUK SSSR SERIYA KHIMICHESKAYA
47. IZVESTIYA VYSSHIKH UCHEBNIKH SERIYA RADIOFIZIKA
48. IZVESTIYA VYSSHIKH UCHEBNIKH ZAVEDENII SERIYA FIZIKA
49. IZVESTIYA VYSSHIKH UCHEBNIKH ZAVEDENII SERIYA KHIMIYA I KHIMCHESKAYA TEKHNOLOGIYA
50. IZVESTIYA SIBIRSKOGO OTIJELENIYA AKADEMII NAUK SSSR SERIYA KHIMICHESKIKH NAUK
51. IZMERITELNAYA TEKHNIKA
52. KARDIOLOGIYA
53. KVANTOVAYA ELEKRONIKA
54. KIBERNETIKA
55. KJNETIKA I KATALIZ
56. KOLLOIDNTI ZHURNAL
57. KOORDINATSIONNAYA KHIMIYA
58. KOSMICHESKAYA BIOLOGIYA I AVIAKOSMICHESKAYA MEDITSINA
59. KRISTALLOGRAFIYA
60. MATEMATICHESKYE ZAMETKI
61. MATEMATICHESKII SBORNJK
62. MIKOLOGIYA I FTIOPATALOGIYA
63. MIKROBIOLOGIYA
64. MOLEKULYARNAYA BIOLOGIYA
65. NEIROFIZIOLOGIYA
66. NEORGANIZHESKYE MATERrALI
67. NEFTEKHIMIYA
68. OKEANOLGIYA
69. OPTIKA I SPEKTROSKOPIYA
70. PARAZITOLOGIYA
71. PISMA V "ASTRONOMICHESKII ZHURNAL"
72. PISMA V "ZHURNAL TEKHNICHESKOI FIZIK"
73. PISMA V "ZHURNAL EKSPERIMENTALNOI I TEORETICHESKOI FIZIKI"
74. POCHVOVEDENIE
75. PRIBORY I TEKHNIKA EKSPERIMENTA
76. PIRKLADNAYA MATEMATIKA I MEKHANIKA
77. RADIOTEKHN1KA I ELEKTRONIKA
78. RADIOKHIMIYA
79. REAKTSIONNAYA SPOSOBNOST ORGANICHESKIKH SOEDINENII
80. SIBIRSKII MATEMATICHESKII ZHURNAL
81. TEORETICHESKAYA I EKSPERIMENTALNAYA KHIMIYA
82. TEORETICHESKAYA I MATEMATICHESKAYA FIZIKA
83. TEORIYA VEROYATNOSTEI I EE PRIMENENIE
84. TEPLOFIZIKA VYSOKIKH TEMPERATUR
85. TERAPEVTICHESKII ARKHIV
86. UKRAINSKII BIOKHIMICHESKII ZHURNAL
87. UKRAINSKII FIZICHESKII ZHURNAL
88. UKRAINSKII KHIMICHESKII ZHURNAL
89. USPEKHI FIZICHESKIKH NAUK
90. USPEKHI KHIMII
91. USPEKHI MATEMATICHESKIKH NAUK
92. FARMAKOLOGIYA I TOKSIKOLOGIYA
93. FIZIKA GORENIYA I VZRYVA
94. FIZIKIA I TEKHNIKA POLUPROVODNIKOV
95. FIZIKA METALLOV I METALLOVEDENIE
96. FIZIKA NIZKIKH TEMPERATUR
97. FIZIKA TVERDOGO TELA
98. FIZIOLOGICHESKII ZHURNAL
99. FIZIOLOGIYA RASTENII
100. FUNKTSIONALNYI ANALIZ I EGO PRILOZHENIE
101. KHIMIKO-FARMATSEVTICHESKII ZHURNAL
102. KHIMICHESKAYA FIZIKA

Notes: List of top Soviet journals reproduced from the International Science Foundation's Individual Emergency Grant application in the Open Society Archives.