Information for replicating results in Kahn and MacGarvie, "How Important is US Location for Research in Science?"

- KM_RS_post.do is the STATA do-file which generates Tables 1-7. This do-file reads in an ascii dataset called KM_RS.out.
- All estimates were obtained using STATA 13.
- Before running the file, list between quotation marks the location of the data and where the output should be stored, e.g.:
    - global DATALOC "c:\research\data"
    - global OUTLOC "c:\research\output"
    - global DOLOC "c:\research\dofiles"
- You will need to install the following STATA add-ins before running the program: eststo, outreg, and ivpois.
- The ivpois_clusterSE.ado file allows for clustering standard errors in ivpois. Save this in the DOLOC directory.
- The variables contained in KM_RS.out are described below. For a more complete description, including the provenance of the data, please see the text and appendix of the paper.

Variable	Description
year	calendar year
fulbrightdummy	1 if fulbright, 0 if control
yearphd	year of PhD completion
region	region of origin
logcgdp	log real GDP per capita of home country
probablegender	gender
top_rank_qs_08	rank of current academic institution
pubcountbyyr	publication count
citcountbyyr	citation count
firstauthcount	count of first-authored publications
firstauthcitc~t	count of first-authored citations
lastauthcount	count of last-authored publications
lastauthcit	count of last-authored citations
highimpcount	count of high-impact publications
highimpcit	count of high-impact citations
pregradpubs	count of publications before PhD graduation
pregradfirstpub	count of first-authored publications before PhD
pregradhiflco	count of high-impact first- or last-authored pubs before PhD
forloc	1 if located outside the US, 0 if in US
lagforloc	forloc lagged 1 year
laglogcurr	log GDP per capita of current location
lagcurreg	lagged current region
rich	1 if high-income home country
laglogcgdp	lagged log real GDP per capita of home country
loghomegdp	Log real gdp per cap of home country 5 yrs prior to grad
highart	1 if high-science home country
fieldnsf	field of study
relrank	relative rank of PhD institution
yearphd2	grouped year of PhD
year2	grouped calendar year
logrank	log of relrank
id2	scientist id
pairid	pair id
expand_cont	propensity score weight
cemfreq	CEM weight
region	region of origin
lagcurreg2	lagged current region for table 1
region2	current region for table 1 (groups Asia and Europe together)
pscore	propensity score