Appendix B: Geography

This appendix describes the geographic dispersion of applicants and analyzes how the grant acts differently in different regions. Using full addresses, I geocoded the locations of each applicant. Figure 1 shows the overall concentrations by Metropolitcan Statistical Area (MSA). The geography of SBIR applicants corresponds to what we would expect of high-tech firms; the largest clusters by far are Boston and San Francisco, and to a lesser extent New York, Los Angeles, Denver/Boulder, and the greater DC metro area. Black (2004) finds concentrations quite similar in his analysis of overall SBIR grants between 1990 and 1995. He also finds that the geographic pattern of SBIR activity is closely correlated to patent geography, which I do not address here.

The number of applicants by award status from the top ten metro regions with the highest number of applicants in 1995 and in 2012 are in Figure 2. San Francisco moves from 9th place to 1st place, reflecting its growing role in the energy technology sector. LA and Boston are near the top of the list in both years. Bridgeport-Stamford and Baltimore fall completely off the list, while NYC and DC enter. This suggests that the changing agglomerations of SBIR winners over time may reflect cities' lifecycles. Graphs by year, not shown here, suggest that the concentration has not changed much over time except for the San Francisco area, which has increased in importance since the mid-1990s.

Wind and solar applicants are in Figure 3, and oil, gas and coal applicants in Figure 4. It is clear that although both renewable and conventional fuel companies locate in major cities, some clustering relates to the area's resource base. Clusters of coal companies in Pittsburgh, PA and oil and gas companies in Houston, TX contrast with clusters of solar firms in Tampa, FL and Orlando, FL.

Figure 5 shows the location of all wind and solar VC portfolio companies in the ThompsonOne database. Agglomeration in Boston and San Francisco, and to a lesser degree in New York, Denver and Chicago are common to both the SBIR applicants (Figure 3) and the VC portfolio companies (Figure 5) in these clean energy sectors. However, the portfolio companies are more concentrated in the major cities where VC firms are also located. We can see this by examining the location of applicants with at least one VC deal (Figure 7), and comparing to the concentration by MSA of all active VC firms in the Preqin database that, according to Preqin, invest in clean technology (Figure 6). In general, the clustering of both VC firms and VC-funded DOE SBIR applicants aligns fairly closely with the clustering of the overall applicant pool. However, there is clearly considerably more clustering of both

VC firms and VC-funded companies in Boston and San Francisco, especially VC-funded companies that have recieved many VC investment rounds.

Meanwhile, there are far more SBIR applicants from Los Angeles and from the greater DC metro area than portfolio companies. For the subset of firms that focus their resources on govenrment grants and procurement contracts, the DC concentration makes sense. LA, somewhat more surprisingly, seems to also be a long-term hub of government-oriented tech companies. For example, Physical Optics - the largest SBIR winner in my data - is located in Torrance, CA, within the LA MSA. The LA government provides supporting activities, such as regular workshops on applying for SBIR grants and other informational and convening resources through its PortTech Los Angeles program, Los Angeles Regional Small Business Development Center, and others.

The visual evidence of agglomeration suggests a test for whether the Phase 1 grant has different effects in specific regions. As in Chen et al. (2010), I create regions that approximate VC investment areas by combining MSAs for greater San Francisco (SF), New York (NY), Los Angeles (LA), Texas triangle (TX), Boston (BOS), and Washington, D.C. (DC).[1] I regress the indicator for subsequent VC investment (VC_i^{Post}) on dummies for each region and interactions of those dummies with the treatment dummy. The primary takeaways from Table 1 are; (a) the grant effect is much higher for firms in SF; (b) outside of SF, the treatment effect does not vary much by city; and (c) in the absence of treatment firms are more likely to receive VC in in TX and unlikely to receive VC in DC. This is consistent with the concentration of oil and gas firms, for whom I find little grant impact, in Houston. The negative coefficient on DC may reflect the preponderance of firms who survive primarily on government grants and do not seek VC finance. On average, DC firms in my data have 50% more previous non-DOE SBIR awards than SF firms.

Specifically, column I of Table 1 includes the full sample, so that the omitted category is all firms not in any of the six regions. The coefficient on treatment alone (that is, tretament when firms are not in any of the six regions) is 9.6 pp, significant at the 1% level, as in the main specifications. The -8 pp coefficient on being in DC, also significant at the 1% level, indicates that losing (untreated) firms in DC are much less likely to recieve VC relative to firms not in the six regions. The 21.4 pp coefficient on the interaction between treatment

[1]The specific MSAs are as follows. 1) SF: San Francisco-Oakland-Fremont; San Jose-Sunnyvale-Santa Clara. 2) NY: New York-Northern New Jersey-Long Island; Hartford-West Hartford-East Hartford; Bridgeport-Stamford-Norwalk; New Haven-Milford. 3) DC: Washington-Arlington-Alexandria,. 4) TX: San Antonio-New Braunfels; Austin-Round Rock-San Marcos; Dallas-Fort Worth-Arlington; Beaumont-Port Arthur; Houston-Sugar Land-Baytown. 5) LA: Los Angeles-Long Beach-Santa Ana; San Diego-Carlsbad-San Marcos. 6) BOS: Boston-Cambridge-Quincy.

and SF indicates that treated firms in SF do much better than untreated firms or firms in other cities. Columns II and III limit the sample to firms in the six regions. In column II, coefficients are relative to being in LA (the omitted region), while in column III no constant term is included to allow estimates for all region dummies. The coefficients on SF alone indicate that relative to the rest of the country, losing applicants from SF are not measurably more likely to receive VC than firms elsewhere.

Across-region results confirm the large SF effect. The regressions in Table 2 limit the sample to situations when a winner is in one region and the loser in another. When the winner is in SF, the treatment effect is more than 25 pp, but when the loser is in SF, the effect is always under 10 pp (I do not show all permutations, but SF's advantage is consistent). Most other combinations suggest treatment effects roughly equal to my main findings. When the winner is in LA and the loser in NY, the effect is 10 pp, and when roles are reversed, it is 9.7 pp. Grants are consistently useful to firms in SF, regardless of whether they are competing with firms locally or far away.

These within- and across-region effects argue against certification as the primary driver. VC firms typically have more information about nearby companies. If certification is driving the grant's effect, there should be less informational content in the grant when competing firms are located in the same MSA and thus a smaller within-region effect. I find no systematically smaller effect within MSAs than between MSAs. Indeed, the effect is highest in the largest VC cluster (SF).

The SF region is obviously a special cluster of companies and investors. In 2012, total VC investment in companies in SF was $10.9 billion, more than the other five regions combined (Florida 2014). Interestingly, Table 1 reveals that relative to the rest of the country, losing applicants from SF are not measurably more likely to receive VC than firms elsewhere. Large knowledge spillovers may help explain the high grant effect in SF. High-tech employees in California, and especially Silicon Valley, exhibit extreme interfirm labor mobility, as suggested by Saxenian (1994) and shown empirically by Fallick, Fleischman and Rebitzer (2006). Rapid job-hopping can increase agglomeration economies, but it imposes costs on employers who must invest in human capital and expose employees to trade secrets. A grant might be more important for startups in a local economy with greater spillovers from R&D investment. The usefulness of the grant in SF may also reflect more intense competition for venture finance. The notion that the benefits of early stage resources are amplified in SF is consistent with Hochberg, Ljungqvist, and Lu's (2007) finding that network benefits for VC performance is twice as large in California as in the whole U.S.

Motivated by the geographic results thus far, I test whether the grant effect is systematically larger in cities with greater VC investment per unit of city output.[2] For example, SF has about \$2 of VC investment for every \$100 of regional output. The next highest region is Boston, at \$0.90, and third is LA, at \$0.22. DC has only \$0.11 for every \$100 of regional output. I use the amount of VC investment per \$100 of MSA output in 2012, for the 20 MSAs with the most VC investment, as a VC intensity index.[3] The left panel of Table 3 interacts the VC intensity index with treatment, and finds a signficant, positive relationship. Below the mean, the grant effect is 7.5 pp, while above it is 17.7 pp. When the regressions are estimated jointly (column VI), the difference is 9.6 pp, significant at the 5% level. This is consistent with the literature. Gans and Stern (2003) conclude that SBIR grantee performance is higher in sectors with high VC capital availability. Lerner (1999) finds that SBIR awardees in the 1980s outperformed a matched set of control firms only in regions with high VC activity.

[2]Calculated using BEA's (2013) statistics on Gross Domestic Product by metropolitan area and Florida's (2014) data on VC investment. A very similar ranking emerges when I use the number of deals, rather than dollar amount.

[3]While I unfortunately do not have data for other years, the ordering of major MSAs by VC and economic activity is likely to be roughly consistent over the time period considered.

Figure 1:

Concentration of Firms in Metropolitan Statistical Areas (All EERE & FE Unique Applicants, 1983-2013)

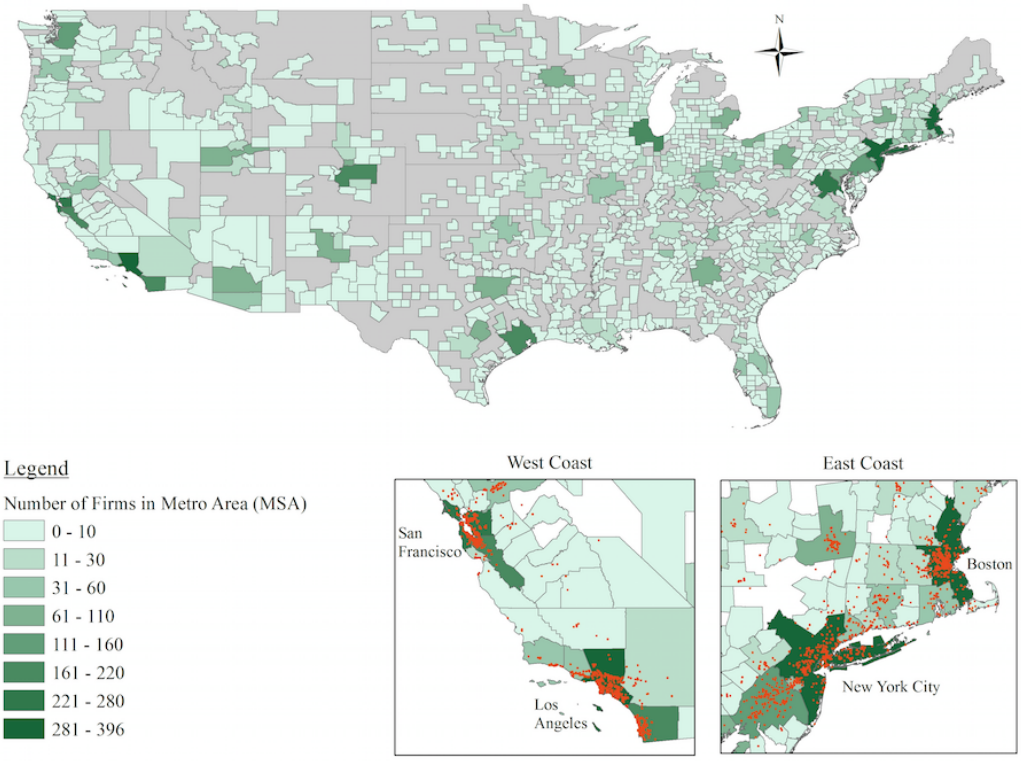

Figure 2:

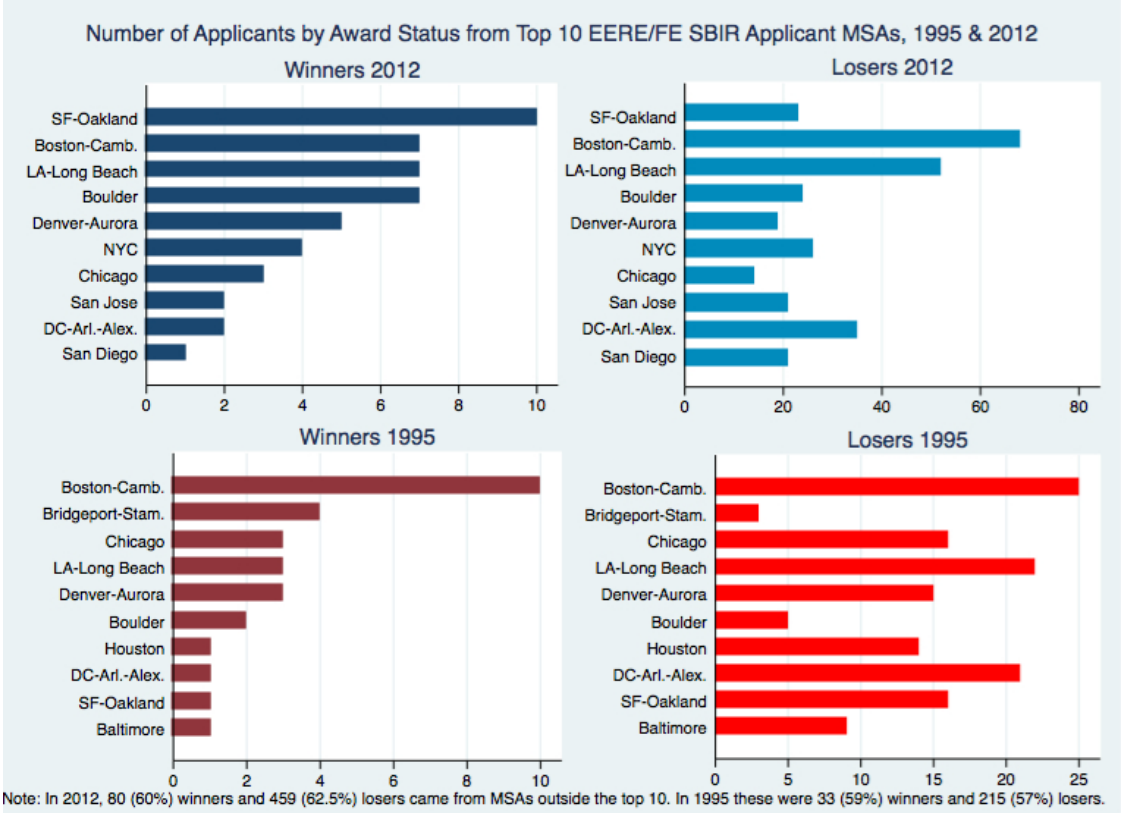

Figure 3:

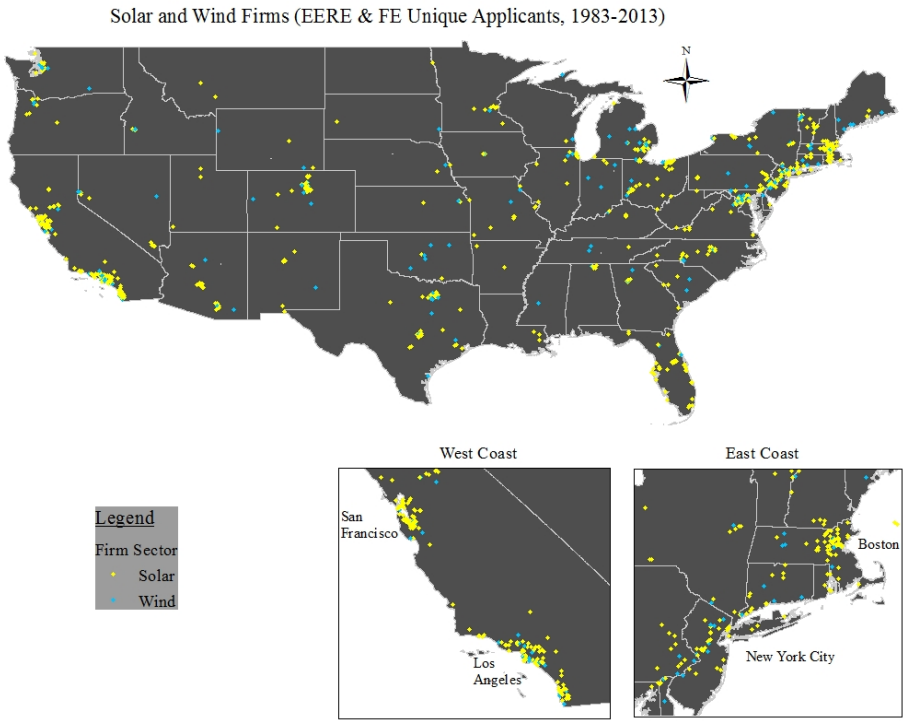

Figure 4:

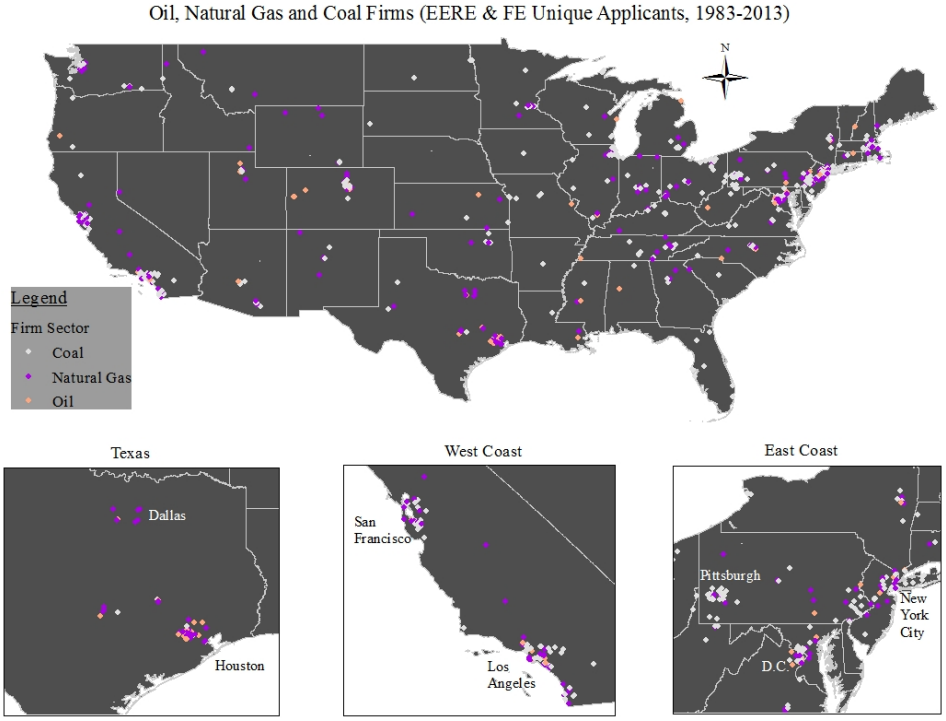

Figure 5:

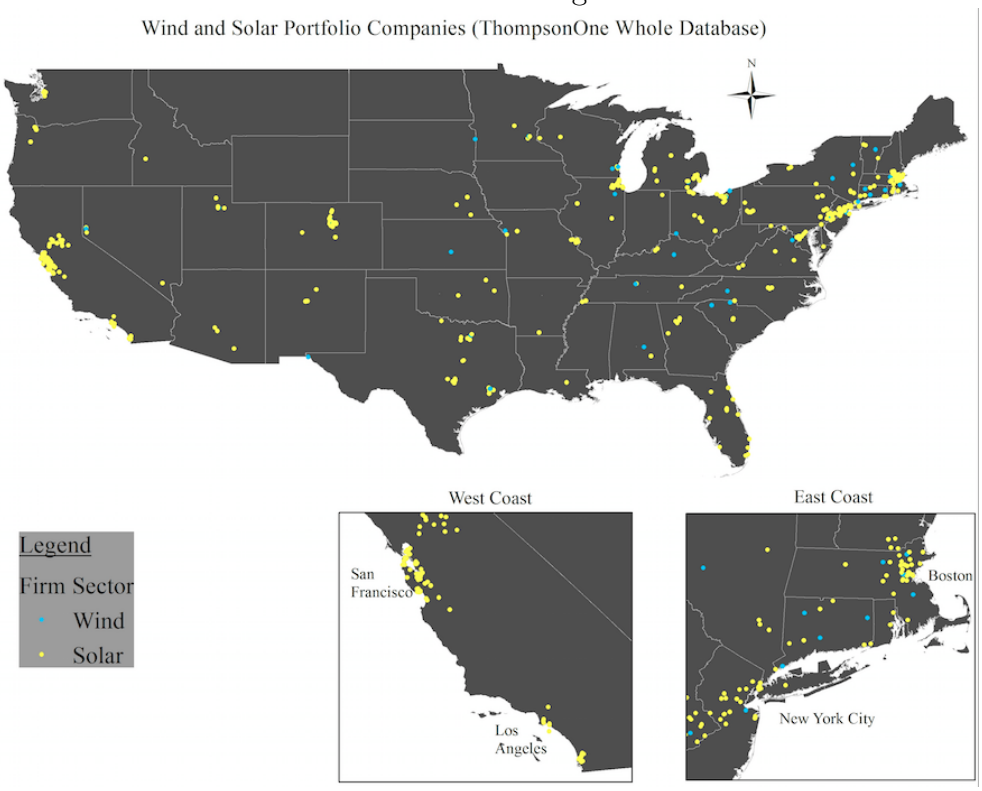

Figure 6:

Concentration of Venture Capital Firms in Metropolitan Statistical Areas (Preqin Whole Database)

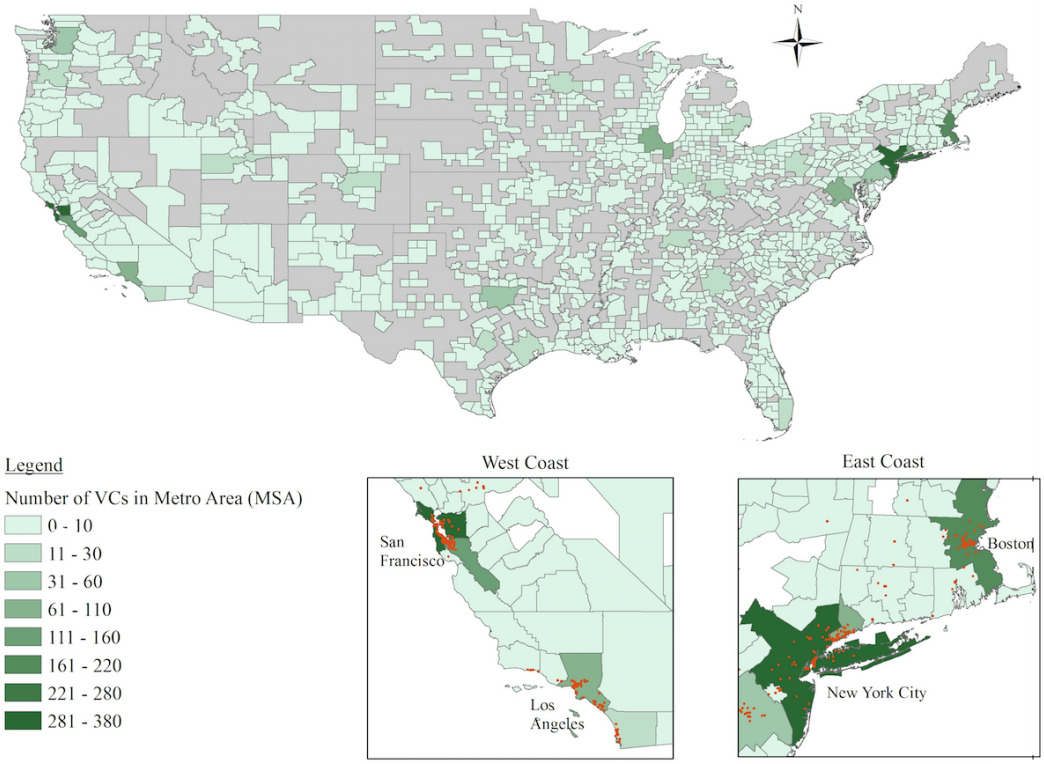

Figure 7:

Firms with at Least 1 Venture Capital Financing Deal (EERE & FE Unique Applicants, 1983-2013)

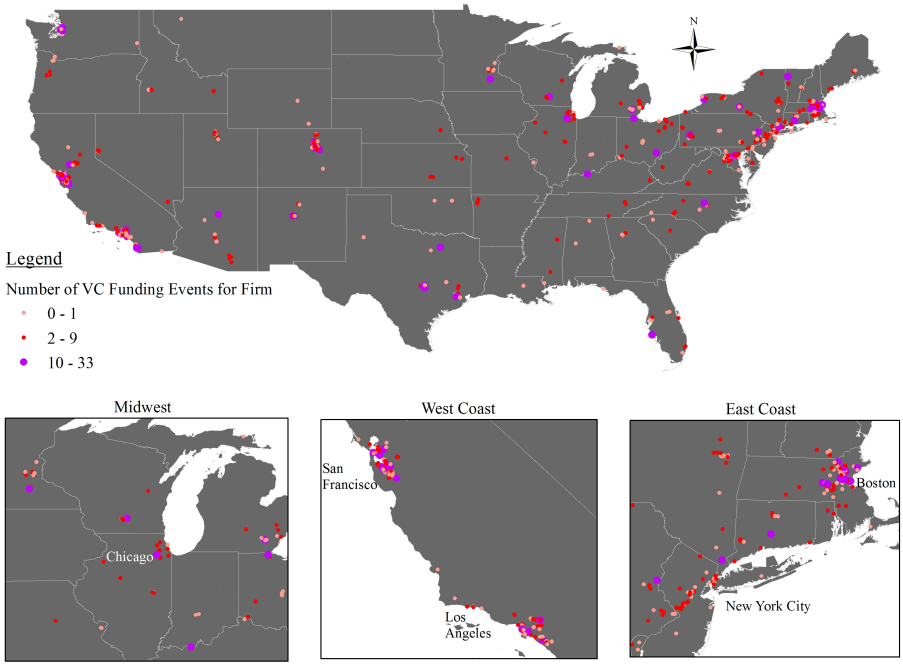

Table 1: Impact of Grant on Subsequent VC for Firms in Six Major Regions

Dependent Variable: VC_i^{Post}	I. Full sample; omitted region is rest of U.S.	II. Sample is 6 regions; no treatment dummy	III. Sample is 6 regions; no constant or treatment dummy
$\mathbf{1} \mid \text{NY}$	0.0285	0.0241	0.197***
	(0.0363)	(0.0615)	(0.0608)
$\mathbf{1} \mid R_i > 0 \cdot (\mathbf{1} \mid \text{NY})$	**-0.0797**	**0.0433**	**0.0433**
	(0.101)	(0.156)	(0.156)
$\mathbf{1} \mid \text{SF}$	0.0307	-0.00717	0.166***
	(0.0296)	(0.0472)	(0.0466)
$\mathbf{1} \mid R_i > 0 \cdot (\mathbf{1} \mid \text{SF})$	**0.214****	**0.438*****	**0.438*****
	(0.109)	(0.156)	(0.156)
$\mathbf{1} \mid \text{BOS}$	0.0129	-0.0301	0.143**
	(0.0404)	(0.0666)	(0.0699)
$\mathbf{1} \mid R_i > 0 \cdot (\mathbf{1} \mid \text{BOS})$	**-0.0132**	**0.139**	**0.139**
	(0.0884)	(0.120)	(0.120)
$\mathbf{1} \mid \text{DC}$	-0.0805***	-0.0963*	0.0771
	(0.0273)	(0.0514)	(0.0516)
$\mathbf{1} \mid R_i > 0 \cdot (\mathbf{1} \mid \text{DC})$	**0.0776**	**0.0761**	**0.0761**
	(0.109)	(0.104)	(0.104)
$\mathbf{1} \mid \text{TX}$	0.0651*	0.0289	0.202***
	(0.0346)	(0.0627)	(0.0584)
$\mathbf{1} \mid R_i > 0 \cdot (\mathbf{1} \mid \text{TX})$	**-0.0734**	**0.0476**	**0.0476**
	(0.0754)	(0.0865)	(0.0865)
$\mathbf{1} \mid \text{LA}$	0.0288	-	0.173***
	(0.0256)		(0.0343)
$\mathbf{1} \mid R_i > 0 \cdot (\mathbf{1} \mid \text{LA})$	**-0.0503**	**0.100**	**0.100**
	(0.0710)	(0.0898)	(0.0898)
$\mathbf{1} \mid R_i > 0$	0.0962***	-	-
	(0.0281)		
VC_i^{Prev}	0.300***	0.286***	0.286***
	(0.0361)	(0.0779)	(0.0779)
$\#SBIR_i^{\text{Prev}}$	0.00107***	0.000932	0.000932
	(0.000262)	(0.000615)	(0.000615)
Competition f.e.	Y	Y	Y
N	3368	1182	1182
R^2	0.353	0.584	0.647

Note: This table is an RD estimating via OLS the impact of the Phase 1 grant ($\mathbf{1} \mid R_i > 0$) on subsequent VC by region (see text for definition), using a bandwidth of 3. In column I all coefficients are relative to the firms outside of the six regions. Columns II and III limit the sample to firms in the six regions, and omit the dummy on treatment. Column III does not include a constant term so that all six regional dummies are estimated. Standard errors robust and clustered at topic-year level. *** $p < .01$. Year$\geq$ 1995

Table 2: Impact of Grant on Subsequent VC for Firms from Different Regions

Dep Var: VC_i^{Post}	I.	II.	III.	IV.	V.	VI.
Winner Region:	SF	NY	SF	BOS	BOS	DC
Loser Region:	NY	SF	BOS	SF	DC	BOS
$\mathbf{1} \mid R_i > 0$	**0.275*** **	**0.0848**	**0.304*** **	**0.0452**	**0.163*** **	**0.0477**
	(0.0754)	(0.0919)	(0.0781)	(0.0663)	(0.0608)	(0.118)
VC_i^{Prev}	0.277***	0.480***	0.248***	0.485***	0.0498	0.208***
	(0.0795)	(0.0881)	(0.0774)	(0.0901)	(0.0627)	(0.0779)
$\#SBIR_i^{\text{Prev}}$	0.000601	0.000306	0.000358	0.000666	0.000403	0.000207
	(0.000674)	(0.000983)	(0.000370)	(0.000794)	(0.000485)	(0.000294)
Topic f.e.	Y	Y	Y	Y	Y	Y
N	325	330	358	346	231	331
R^2	0.215	0.250	0.173	0.253	0.173	0.105
	VII.	VIII.	IX.	X.	XI.	XII
Winner Region:	NY	BOS	LA	NY	TX	DC
Loser Region:	BOS	NY	NY	LA	DC	TX
$\mathbf{1} \mid R_i > 0$	**0.0774**	**0.0314**	**0.102***	**0.0972**	**0.109***	**0.168**
	(0.0846)	(0.0675)	(0.0584)	(0.0802)	(0.0602)	(0.115)
VC_i^{Prev}	0.187**	0.250***	0.294***	0.233***	0.203**	0.0815
	(0.0776)	(0.0816)	(0.0807)	(0.0821)	(0.0807)	(0.150)
$\#SBIR_i^{\text{Prev}}$	0.000406	0.000932	0.00135**	0.00296***	-0.000107	-0.000203
	(0.000349)	(0.000700)	(0.000568)	(0.000850)	(0.000139)	(0.000327)
Topic f.e.	Y	Y	Y	Y	Y	Y
N	346	329	333	493	213	267
R^2	0.108	0.130	0.178	0.152	0.209	0.084

Note: This table is an RD estimating via OLS the impact of the Phase 1 grant ($\mathbf{1} \mid R_i > 0$) on subsequent VC by location, where a winner is in one region and a loser in another. These regressions use BW=all, but usually there are only 2 firms per topic. Standard errors robust and clustered at topic-year level. *** $p < .01$. Year$\geq$ 1995

Table 3: Impact of Grant on Subsequent Venture Capital by Regional VC Intensity

Dependent Variable: VC_i^{Post}	Treatment interacted with regional VC intensity			Separate regressions around mean regional VC intensity		
	I. BW=2	II. BW=3	III. BW=all	IV. $VC\ Int \leq 6.2$	V. $VC\ Int > 6.2$	VI. Diff $VC\ Int \leq 6.2$ & > 6.2
$(\mathbf{1} \mid R_i > 0) \cdot VC\ Int$	**0.0157****	**0.0159*****	**0.0112****			
	(0.00791)	(0.00602)	(0.00480)			
$\mathbf{1} \mid R_i > 0$	0.118**	0.120***	0.0979***	**0.0750****	**0.177*****	0.171***
	(0.0464)	(0.0402)	(0.0343)	(0.0356)	(0.0522)	(0.0343)
$\mathbf{1} \mid R_i > 0 \cdot (\mathbf{1} \mid VC\ Int \leq 6.2)$						**-0.0960****
						(0.0456)
VC_i^{Prev}	0.329***	0.322***	0.330***	0.336***	0.345***	0.347***
	(0.0734)	(0.0623)	(0.0444)	(0.0509)	(0.0593)	(0.0341)
$\#SBIR_i^{\text{Prev}}$	0.000341	0.000591	0.000602*	0.00101*	0.000287	0.000326
	(0.000652)	(0.000531)	(0.000334)	(0.000517)	(0.000434)	(0.000238)
$VC\ Int$	0.00000360	-0.000367	0.00178			
	(0.00222)	(0.00165)	(0.00129)			
Competition f.e.	Y	Y	Y	N	N	N
Topic f.e.	N	N	N	Y	Y	Y
Topic f.e. $\cdot (\mathbf{1} \mid VC\ Int \leq 6.2)$	N	N	N	Y	Y	Y
N	1290	1538	2571	1522	1050	2571
R^2	0.536	0.498	0.371	0.286	0.356	0.327

Note: This table is an RD estimating via OLS the impact of the Phase 1 grant ($\mathbf{1} \mid R_i > 0$) on subsequent VC by regional VC intensity. VC intensity is calculated as 2012 ($ VC investment)/(100$ of GDP) for each of the twenty MSAs with the most VC investment. Each column includes only data from firms in the relevant range of VC intensity. In the difference regression (column VI), all covariates are interacted with a dummy for low VC intensity. Topic-level dummies are used to have adequate firms per fixed effect. Coefficients on $VC_i^{\text{Prev}} \cdot (\mathbf{1} \mid VC\ Int \leq 6.2)$, $\#SBIR_i^{\text{Prev}} \cdot (\mathbf{1} \mid VC\ Int \leq 6.2)$ and $(\mathbf{1} \mid VC\ Int \leq 6.2)$ not reported for space concerns. Standard errors robust and clustered at topic-year level. *** $p < .01$. Year$\geq$ 1995