Research Spillovers, Two Ways

by Kyle Myers & Lauren Lanahan

Discussion by Douglas Hanley (Pitt)

REER 2019, Georgia Tech
doughanley.com/slides/reer_2019

Great paper!

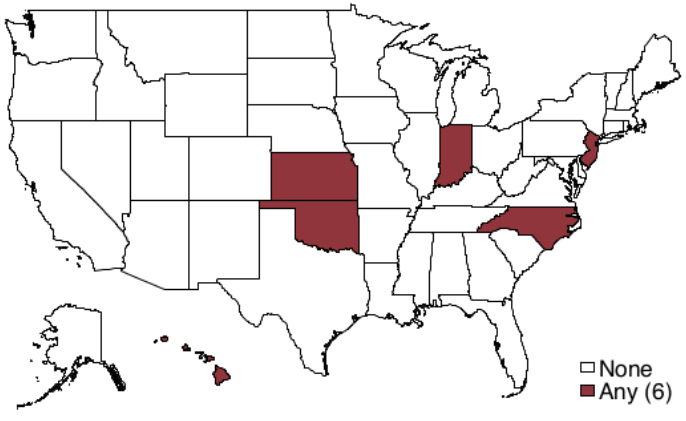
Excellent use of state-of-the-art text analysis methods to construct match between grant topics and technology classes

Critically important question in research and innovation policy: what is the nature of spillovers across field and across space?

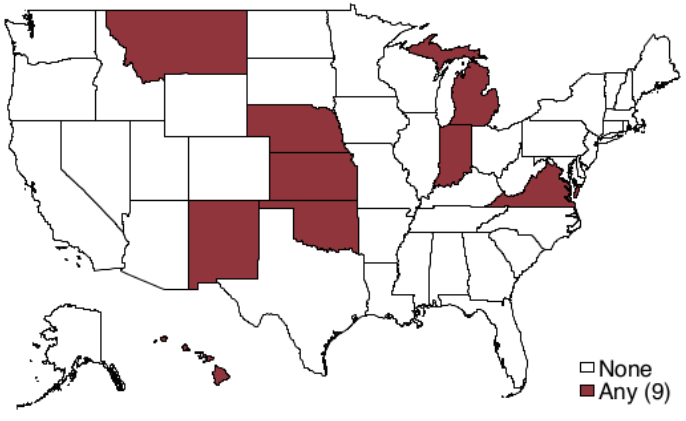
Clever identification strategy to generate exogenous variation in field-level research funding.

Source of randomization

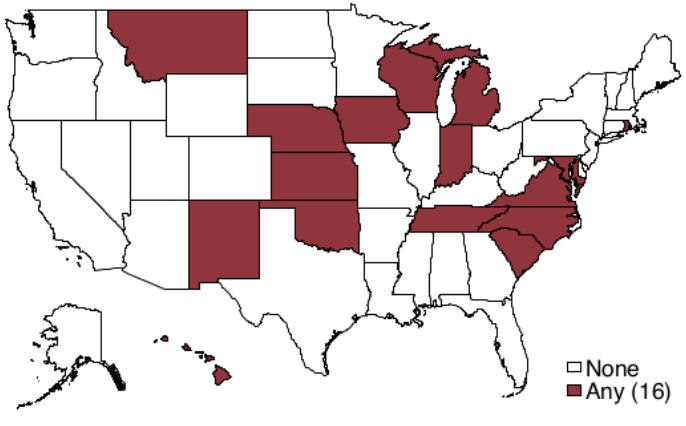
(a) 2006



(b) 2012



(c) 2017



(d) Match Amounts

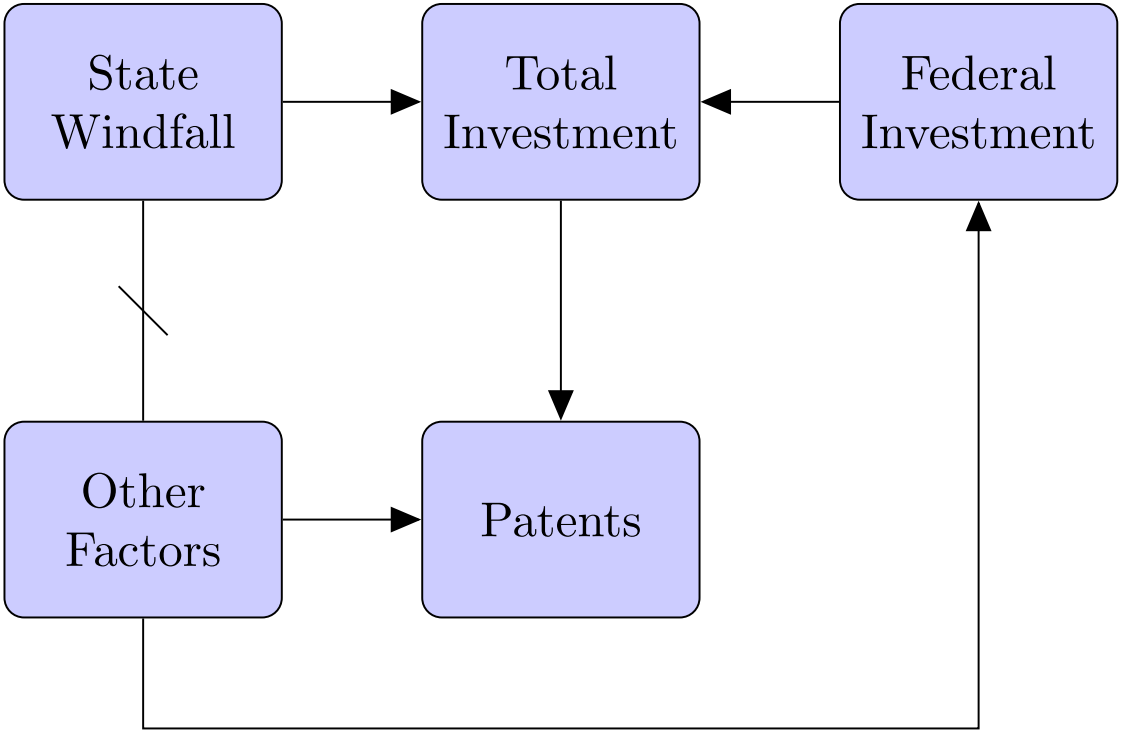
Phase I

Amount, $K	N State-Years
25-45	27
50-65	46
75	31
100-105	14

Phase II

Amount, $K	N State-Years
50-75	10
100	17
250	10
500	24

Identification strategy



Gotta love those causal graphs

Methods

Distributing Spillovers

Funding from FOA categories gets distributed to technology classes using "spillover radius" $\bar{s}$.

$$\Big[\mathbf{s_1} \quad \mathbf{s_2} \quad \mathbf{s_3} \quad \mathbf{s_4} \quad \mathbf{s_5} \quad \mathbf{s_6} \underbrace{\Big|}_{\bar{s}=6} \quad s_7 \quad s_8 \quad s_9 \quad s_{10}\Big]$$

Classes ranked in top $\bar{s}$ in terms of similarity *ranking* get even split of FOA funds.

- This has the effect of imposing the same spillover radius for each FOA category.

Using intensity information

The text similarity matrix (between FOA topic and tech class) contains lots of rich information

- could set an minimum absolute threshhold $\underline{s}$ (or not)
- FOA funds allocated to tech classes based on relative similarity

Could add in some additional text sources

- actual CPC description text
- upweight title and category info

Text categorization

If CPC isn't designed for research, why are we using it? One option would be to make your own groups

- Could K-means cluster document vectors for patents into similar groups.
- This would naturally be conducive to the FOA classification.

Same method could be used to cluster FOAs into a new topic categorization and/or used to validate FOA topic use

Constructing validation set

Always great (and fun!) to construct a validation set to assess quality of match

- Is it always crystal clear to a human what the correct categorization is?
- Is there always exactly one tech class that a particular FOA topic matches to?
- How does the algorithm itself perform relative to human assessment?

Construction of windfall instrument

Might be some issues with the intercept and independence between windfall and federal funding.

$$\operatorname{arcsinh}(K_{jt}^{\text{loc}}) = \Gamma \log(K_{jt}^{\text{fed}}) + \mu_{jt}$$

Would be great to have intercept so $\mathbb{E}[\log(K_{jt}^{\text{fed}})\mu_{jt}] = 0$

Might be even better to reformulate in linear terms, at least for instrument

$$K_{jt}^{\text{loc}} = \alpha + \beta K_{jt}^{\text{fed}} + \mu_{jt}$$

Specification

Substitution between private and public

Fundamentally, we need to be aware here that with these grants, we're merely adding funding to existing operations.

- These are small businesses, but they already do lots of research presumably, and money is pretty fungible in this case.

There are conceptual issues here relating to how firms internal investment responds to grant receipt, which would affect the cost estimates.

- Could go up or down, they might collaborate with university or lab

Existing private investment

Thinking about Cobb-Douglas and baseline investment levels, let B_j be baseline private investment and K_j be the public grant

$$\log(y) = \alpha \log(B_1 + K_1) + \beta \log(B_2 + K_2)$$

The marginal effect of the studied regression would be

$$\frac{\partial \log(y)}{\partial \log(K_1)} = \alpha \left(\frac{K_1}{B_1 + K_1} \right)$$

If funding is proportional, so that $K_1 = \gamma_1 B_1$, we get

$$\frac{\partial \log(y)}{\partial \log(X_1)} = \alpha \left(\frac{\gamma_1}{1 + \gamma_1} \right)$$

Data on private investment

Would it be possible to bring in, at least expositionally, some of the BRDS data from NSF? You'd need a NAICS/CPC crosswalk, which might exist, but could also be done using patents/Compustat.

Then at least you could control for something like the average γ per class.

- Or just class fixed effects?

Interpretation

Interpretation of results

We're not really getting full costs, because we don't see the private investment response of the firms.

Particularly with financial and cash flow considerations, this might be important?

- This is true in a static setting, but probably even more important in a dynamic setting, where firms get multiple rounds of grants over years.

Reallocation of scientists

At the end of the day, we know that the stock of scientists is a pretty slow moving object. In the short-run, rather fixed

The question then is how much of the effects that we see here are due to reallocation of scientists/effort from one area of research to another?

Depends on how easy it is for scientists to switch areas.

- The more difficult it is, the more plausible these results seems taken at face value.

Big implications

The 2D graphs very cool! But they're a little hard to interpret sometimes

- I would expand on some of the back of the envelope cost estimates to complement these

Would also be great to see more investigation of potential mechanisms

- For example, how does this all interact with the national labs?

Thanks!