

Question 1: Young Farmers of America

In recent years, the American agricultural sector has undertaken substantial consolidation. A relatively small number of large farm corporations produce the majority of agricultural output. These changes have generated massive increases in productivity. But they also mean that there are far fewer “farmers” than there were in the past.

Some analysts and advocates argue that that farming has a steep “learning curve” that is hard to pursue without essentially growing up on a farm. This may be one reason why we don’t see more new entrants and early career farmers in the US. One pathway is for early stage farmers to lease land from established farmers. By farming a piece of land alongside a more experienced mentor, young farmers may develop the skills they need to become self-sufficient. In fact, many established farmers do own land that they do not themselves cultivate. And it is not uncommon for farmers to lease their unused land to other farmers for specified periods of time.

However, leases are rarely made to “early career farmers”. One reason may be that the early career farmers are risky bets. There is a good chance that an early career farmer will not earn enough revenue from farming to pay the costs of the lease, let alone living expenses. Thus, it is less risky for established farmers to lease the land to an established operation for whom ability to pay and expected revenue is more dependable.

In an effort to increase the number of small farm owning households, about five states have begun experimenting with subsidies, generally referred to as “Beginning Farmer Tax Credits” (BFTC). Established farmers can claim these tax credits when they lease a portion of their farm or ranch land to a “new” farmer. The BFTC is intended to give established farmers an incentive to take on the extra financial risk (and hopefully provide mentoring services) associated with giving the lease contract to the new farmer.

Part 1

- (a) Develop an economic model to help understand how a state BFTC might affect the prevalence of new farmers in the state. What factors will determine how effective the BFTC is apt to be in the real world? Make sure to give precise explanations of the inputs to the model, what is being maximized by the agents in the model, and what tradeoffs are at play in the model.

Part 2

It turns out Iowa implemented a BFTC in 2015. “Existing farmers” apply for the BFTC when they lease to a “beginning farmer”. Applications for the BFTC are reviewed by the BFTC committee. The statute is vague on who counts as a “beginning farmer” so the committee uses its discretion on this dimension of eligibility. They also limit eligibility to cases where the “beginning farmer” has less than \$600,000 in net wealth.

The state has the following tax data on all farmers producing in the state since 1995 to current:

- i. Gross Farm Sales
- ii. Income from Farming Activity
- iii. Total Income
- iv. Acres of Farmland Owned
- v. Acres of Farmland Leased from another operator
- vi. Acres of Farmland Leased to another operator
- vii. Tax Credits Claimed (Including BFTC)
- viii. Farm Property Address
- ix. Capital Expenses
- x. Debt Expenses

Naturally, a “beginning farmer” will only appear in the administrative tax data if they succeed in getting farmland (leased from an existing operator or not) and filing the associated tax forms. For the most part, you can obtain this type of tax data in other states as well.

- (a) Using the model you developed in part 1, generate hypotheses about the effect of Iowa's BFTC program that the state should be interested in. For each hypothesis, describe how it follows from the intuition of your theoretical model. Be sure to express in the hypothesis statement whether the effect of the policy is positive, negative, or ambiguous. (Your model might suggest countervailing effects that results in ambiguous predictions. This is fine and you can simply point it out.) To the extent possible, you should describe indicators you could construct from the data provided that might help you test the hypothesis and comment on how well the data-based indicators connect with the theory.
- (b) Identify some hypotheses that are interesting but do not emerge directly from your model, or are not testable from the available data. Comment what is missing/ignored in the theoretical model or data that would have to be changed in order to explore the new hypothesis. If possible, describe an existing literature where the new hypothesis, modelling framework, or type of data would be at home.
- (c) Other states are interested in adopting a policy similar to Iowa's. To what extent do you think empirical evidence from Iowa's policy will be useful for other states? Is there anything from your model that will inform your point of view?
- (d) The University of Iowa is a land grant institution that wishes to hire a new assistant professor whose research agenda will be attractive to policy makers in the state. Develop a high level research plan around the BFTC in Iowa that will result in publishable academic work. Your research plan can involve causal program evaluation, exploration of policy relevant economic parameters in the agricultural sector, or a predictive model for describing policy impacts, such as Computable General Equilibrium model. Either way, you should clearly describe the key details of your overall plan: research question, unit of analysis, research design, statistical model, assumptions, threats to validity, and proposed sensitivity checks).

Question 2: Ohio's Vax-A-Million Program

The Ohio Vax-A-Million program was a weekly lottery held every Wednesday from 26 May 2021 to 23 June 2021 by the Ohio Department of Health in conjunction with the Ohio Lottery. Every Ohio resident who had been vaccinated with at least one dose of the Covid-19 vaccine as of the Saturday before the lottery was eligible to win the lottery. In each lottery, one person's name would be drawn and the prize was \$1 million.

Part 1

- a) Develop some mathematical notation to precisely describe what it would mean for the Vax-A-Million program to “generate” more vaccinations.
- b) Elena -- an analyst from the Ohio Department of Health -- contacts you to ask your views on how many vaccinations the Vax-A-Million program would need to “generate” in order to pass a benefit cost test? She notes that (i) the Ohio Government always uses a Value of Statistical Life of \$11 million when it does Cost-Benefit studies, (ii) the vaccine trials seem to imply that the vaccine reduced “mortality rates” by around 90%, (iii) about 13,000 people died of Covid-19 in Ohio in 2020, and (iv) the population of Ohio is about 11.78 million. Elena feels like she has compiled most of the relevant information but she isn't sure how to make a back of the envelope calculation of the number of vaccinations that would seem to justify the program on benefit-cost grounds. Write a short paragraph explaining how you would try to answer her question. Any back of the envelope calculation involves assumptions. Which assumptions are most important in your calculation?
- c) Is there a good conceptual case for a government program like the Vax-A-Million lottery? What rationales for government policy seem to be relevant considerations in this situation? Is this about paternalism? Adverse selection? Redistribution?
- d) Elena thinks it's pretty clear that the program worked. She says that her office did a review of the data and found that on May 1 about 4.89 million people had been vaccinated in Ohio. By June 30, when the program was wrapping up, the cumulative number of people vaccinated was up to about 5.66 million. “That's over three quarters of a million people we helped get vaccinated!”, she says in a proud voice.

Expand the notation you developed in part (a) to describe Elena's estimate of the effect of Vax-A-Million program. Under what assumptions will her estimator be unbiased/consistent for the causal effect of the program on vaccinations? Are these assumptions plausible? Are they testable? Why might the assumptions fail in practice?

Part 2

It turns out that several states adopted lotteries like Ohio's Vax-A-Million program. The table below shows the date of the lottery announcement in each lottery state.

State	Date of Lottery Announcement	Maximum Cash Prize
Ohio	5/12/2021	\$1,000,000
Delaware	5/25/2021	\$302,000
California	5/27/2021	\$1,500,000
New Mexico	6/1/2021	\$5,000,000
West Virginia	6/1/2021	\$1,000,000
Arkansas	6/1/2021	\$1,000,000
Washington	6/3/2021	\$1,000,000
Kentucky	6/4/2021	\$1,000,000
North Carolina	6/10/2021	\$1,000,000
Massachusetts	6/15/2021	\$1,000,000
Maine	6/16/2021	\$896,809
Illinois	6/17/2021	\$1,000,000
Louisiana	6/17/2021	\$1,000,000
Nevada	6/17/2021	\$1,000,000
Michigan	7/1/2021	\$2,000,000

Mark is thinking about how to estimate the effects of lottery programs on vaccine take up. He starts by defining V_{st} to represent the total (cumulative) number of people in state s who have ever been vaccinated as of date t . He lets L_{st} be a dummy variable indicating that a lottery program has been announced in state s as of date t . He collects data on V_{st} from every state in the country, and uses the table above to determine the value of L_{st} each state and date. The states that don't have a lottery program always have $L_{st} = 0$. Mark starts with the following regression:

$$V_{st} = \beta L_{st} + a_s + b_t + \epsilon_{st}$$

- Mark points out that a_s and b_t are unmeasured variables that are probably endogenous. But he thinks he should be able to estimate β by fitting a simple OLS regression on "within" transformed outcomes and lottery variables. Show how to transform the data. And write down the estimator Mark has in mind.
- Under what assumptions will Mark's estimator be consistent for the causal effect of the lottery program on vaccination take up? Discuss both the technical and substantive implications of each assumption and suggest how the assumption might be tested or evaluated or probed in applied work.
- Anil says that he has always thought that estimates from models like Mark's are best understood as a kind of average causal effect of the programs adopted in several different states. How has that view changed because of recent work in the literature on two way fixed effects regressions and difference-in-difference designs?
- Marcel points out that lottery treatment effects might be heterogenous across states or across time periods. Why might treatment effects differ across states or over time? Which problem do you think is more problematic for the "approximating" model that Mark proposed above?
- Choose a single Lottery State and explain how you could implement 2 x 2 difference in difference to estimate of the effect of the lottery program. What assumptions would you make in order for your estimator to uncover a well-defined causal effect? What causal parameter is this 2x2 DID estimating? Can this

approach be used to uncover effects at different points in time as well? Would statistical inference be a problem in this 2 x 2 setting?

- f) How would you proceed with this project? Is it practical to simply estimate individual 2 x 2 DID parameters for each treated state? Are there good ways to summarize the heterogeneity?
- g) Vinny says that the new DID literature is too confusing. He has heard that synthetic control methods are superior in any case. However, he is not really sure why people say that? “They both seem like versions of a comparative interrupted time series to me,” he says.

Explain how you could use a synthetic control approach to build an estimator of the causal effect of a lottery program on vaccine take up. Under what assumptions would the synthetic control estimator perform well? Are the assumptions credible in this application? How do they differ from the two way fixed effects approach described above? Sketch out the basic steps you would take to implement a synthetic control study of the effects of a lottery program.

- h) While working on the project, you learn that on 10 May 2021, the US announced expanded vaccine eligibility for 12 to 15 year olds. How should this information change your analyses of (i) the original model proposed by Mark, (ii) the modern (heterogenous effects) DID analysis, and (iii) the synthetic control approach?