Farmers Markets and Food-Borne Illness*

Marc F. Bellemare† *Ngoc (Jenny) Nguyen*‡

February 5, 2017

Abstract

Using administrative longitudinal data on all US states and the District of Columbia for the years 2004, 2006, and 2008-2013, we study the relationship between farmers markets and food-borne illness. We find a positive relationship between the number of farmers markets per million individuals and the number of reported (i) total outbreaks and cases of food-borne illness, (ii) outbreaks and cases of norovirus, and (iii) outbreaks of campylobacter per million in a given state-year. When we exploit weather shocks as a source of plausibly exogenous variation for the number of farmers markets per million, the majority of the aforementioned positive relationships persist. Allowing for small departures from the assumption of strict exogeneity of weather shocks, the relationship between farmers markets per million and the number of reported (i) total cases of food-borne illness as well as (ii) cases of norovirus per million turn out to be robust. Our estimates indicate that for every additional farmers market per million, there are six additional cases of food-borne illness per million, and that a doubling of the number of farmers markets in the average state-year would be associated with an economic cost of at least $220,000. Our core results are robust to different specifications and estimators as well as to deleting outliers and leverage points, and falsification and placebo tests indicate that they are unlikely to be spurious.

Key words: Food Safety, Food-Borne Illness, Farmers Markets

JEL Classification Codes: I12, I18, Q13

* We thank the University of Minnesota's Healthy Foods, Healthy Lives Institute for financial support. We are also grateful to Ed Ragland from the USDA's Agricultural Marketing Service for sharing the farmers markets data, to Rob King for extensive comments on earlier versions of this paper, to Jayson Lusk for his generous help clarifying the question of reverse causality via his monthly FooDs survey, as well as to Wolfram Schlenker and Jesse Tack for respectively sharing and helping with the weather data. We also thank Tim Beatty, Francisco Diez-Gonzalez, Joe Ritter, and Dawn Thilmany as well as seminar participants at the University of Georgia, Ohio State University, and Oklahoma State University and conference participants at the 2015 International Conference of Agricultural Economists in Milan for comments and suggestions which led to a substantially improved manuscript. Finally, we are grateful to those readers who took time to write with constructive criticism in response to the January 2016 *New York Times* Gray Matter that discussed preliminary results for this article. All remaining errors are ours.
† Corresponding Author, Associate Professor, Department of Applied Economics and Director, Center for International Food and Agricultural Policy, University of Minnesota, 1994 Buford Avenue, Saint Paul, MN 55108, Email: mbellema@umn.edu.
‡ PhD Student, Stephen M. Ross School of Business, University of Michigan, 701 Tappan Ave, Ann Arbor, MI 48109, Email: jennyn@umich.edu

1. Introduction

Since the mid-1990s, the number of farmers markets in the United States has grown almost fivefold, rising steadily from 1,755 to 8,268 farmers markets between 1994 and 2014 (USDA Agricultural Marketing Service, 2015).

Given that farmers markets often sell foods from producers who are subject to a less stringent set of regulations than the foods sold at convenience stores, grocery stores, super markets, and big-box stores, what does the recent rise in popularity of farmers markets mean for food-borne illness? Is the number of farmers markets in a given state associated with food-borne illness in the same state in any systematic way, if at all?

On this, there is little to no empirical evidence. Sivapalasingam et al. (2004) do not specifically look at foods sold at farmers markets, but they document how fresh produce has been a growing cause of food-borne illness outbreaks (i.e., *Salmonella*, *Cyclospora*, and *E. coli*) in the United States. Francis et al. (1999) note that minimally processed vegetables provide new ecosystems within which pathogens can emerge and evolve. Only a handful of studies specifically look at food sold at farmers markets. The first such study is by Park and Sanders (1992), who analyze over 1,500 samples of 10 different types of vegetables at farmers markets and supermarkets and find that vegetables from farmers markets are much more likely to contain *Campylobacter*, and thus much more likely to pose health hazards. Another is a recent study by Scheinberg et al. (2013), who find that chicken sold at farmers markets is more likely to test positive for *Salmonella* or *Campylobacter* than chicken sold at supermarkets. Relatedly, Harrison et al. (2013) interview 45 farmers market managers in Georgia, Virginia, and South Carolina about their food safety practices and find that "over 42% [of farmers market managers] have no food safety standards in place for [their] market … less than 25% of managers sanitize market surfaces [and] … [o]ver 75% of markets offer no sanitation training to workers or vendors." In addition to those three studies, a recent report discussing the results of a nationally representative monthly survey of food demand notes that "… people who shopped or ate at farmers markets [in the past two weeks] were much more likely (20% vs. 2.5%) to say they had food poisoning in the past two weeks than people who did not eat or shop at a farmers market" (Lusk, 2016).

On the one hand, the remotely produced and procured foods typically sold at convenience stores, grocery stores, supermarkets, and big-box stores are produced in the context of long agricultural value chains by large, often multinational firms that face serious scrutiny from food-safety authorities. Those firms face serious incentives to apply the strictest possible food-safety protocols, which would lead one to believe that remotely produced foods could lead to fewer, albeit more widespread, outbreaks of food-borne illness. Locally produced and procured foods typically sold at farmers markets, on the other hand, travel much shorter distances, they are handled by fewer people, and they are generally consumed more quickly. This would lead one to believe that locally produced foods could lead to fewer albeit more geographically concentrated outbreaks of food-borne illness. There is thus no *a priori* reason to believe there is any systematic relationship between farmers markets and food-borne illness. And even if there were such a relationship, it is not immediately obvious whether it would be positive or negative.

We study the relationship between farmers markets and food-borne illness. Specifically, we look at the relationship between the number of farmers markets per million individuals in a given state in a given year on the one hand and, on the other hand, the number of reported outbreaks of food-borne illness and cases of food-borne illness per million.[1] We do this for the total reported number of

[1] Because outbreaks encompass several cases, we differentiate between outbreaks and cases of food-borne illness throughout this paper. Looking separately at outbreaks and cases allows testing that our results are robust to different specifications of the dependent variables. For the sake of brevity, we will sometimes talk in this paper of "outbreaks of food-borne illness" and "cases of food-borne illness" to refer to all *reported* outbreaks or cases of food-borne illness. When referring to specific illnesses rather than to total numbers of reported outbreaks or cases, we will refer to them by name (e.g., norovirus, *Campylobacter*, etc.). It should also be implicit throughout this paper that we only refer to *reported* outbreaks or cases of food-borne illness, since those are the only ones available in the data.

outbreaks and cases of food-borne illness per million no matter the type, but also for reported outbreaks and cases of the seven most common (in terms of outbreaks) illnesses in the data, i.e., norovirus, *Salmonella enterica*, *E. coli shiga*, *C. perfringens*, *Campylobacter jejuni*, the scombroid toxin, and *Staphylococcus aureus*. To do so, we begin by exploiting variation over time and space in a state-level administrative panel data set covering all 50 US states and the District of Columbia for the period 2004, 2006, and 2008-2013. In addition to the *a priori* ambiguous relationship between farmers markets and food-borne illness discussed above, there is no reason to believe that, should there be a systematic relationship between farmers markets and food-borne illness, it would show up as statistically significant when looking at such an aggregate level. Consequently, finding any statistically significant relationship in this context would constitute *prima facie* evidence in favor of the existence of a potentially causal relationship between farmers markets and food-borne illness.

We then successively include (i) the number of farmers markets in neighboring states, in order to ensure that our results are not driven by a violation of the stable unit treatment value assumption (SUTVA; cf. Pearl, 2009) operating through spillovers, (ii) a linear time trend, (iii) state-specific linear time trends, and finally (iv) US Census Bureau regional division-year fixed effects,[2] all in an effort to ensure that our results are robust to different specifications. We further report estimation results for semiparametric (i.e., splines, to account for the potentially nonlinear relationship between farmers markets and food-borne illness) robustness checks. Lastly, placebo and falsification tests suggest that our core results are unlikely to be spurious.

[2] The US Census Bureau divides the United States in four regions (Northeast, Midwest, South, and West), which are themselves divided into regional divisions. The Northeast includes New England (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont) and the Mid-Atlantic (New Jersey, New York, and Pennsylvania); the Midwest includes the East North Central (Illinois, Indiana, Michigan, Ohio, and Wisconsin) and West North Central (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota); the South includes the South Atlantic (Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, the District of Columbia, and West Virginia), and the East South Central (Alabama, Kentucky, Mississippi, and Tennessee), the West South Central (Arkansas, Louisiana, Oklahoma, and Texas); and the West includes the Mountains (Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming) and the Pacific (Alaska, California, Hawaii, Oregon, and Washington).

In an effort to make a causal statement, we then exploit the variation in weather between states in a given year as a source of plausibly exogenous variation in the number of farmers markets per million. Specifically, we use the variation in average minimum daily temperatures in each state-year, the logic being that, controlling for a linear time trend, this measures unpredictable within-state shocks to average minimum daily temperature, and that for negative or positive such shocks, a state is less or more likely to see more farmers markets open in a given year, respectively. Lastly, using Conley et al.'s (2012) methodology, we explore the robustness of our estimates to small departures form the assumption of strict exogeneity for our instrumental variable.

Ultimately, we find a positive, statistically significant, robust relationship—one that appears to be causal if one believes in the validity of our instrumental variable—between the number of farmers markets per million and the number of reported (i) total cases of food-borne illness and (ii) cases of norovirus—the most common cause of viral gastro-enteritis, which causes vomiting and diarrhea, and which kills over 570 people annually in the US (US Centers for Disease Control and Prevention, 2016a)—per million. Just as importantly, we find no statistically significant relationship between farmers markets and some common types of food-borne illness, viz. *Salmonella enterica*, *E. coli shiga*, *C. perfringens*, scombroid food poisoning, and *Staphylococcus aureus*. In other words, the presence of farmers markets in the average state-year appears to cause some but not all common types of food-borne illness in the United States for the period 2004, 2006, and 2008-2013.

From an economic perspective, this matters because food-borne illness is estimated to cost $51 billion annually in the United States, with an average cost of $1,068 per case of food-borne illness (Scharff, 2012).[3] From a public health perspective, this matters because food-borne illness causes over 55,000 hospitalizations and almost 1,400 deaths annually in the United States (US Centers for Disease

[3] These are Scharff's more conservative estimates. The estimates from what he dubs his "enhanced" model are for a total cost of food-borne illness of $77 billion, and an average cost per case of $1,626.

Control and Prevention, 2014). Using Scharff's $1,068-per-case figure, our back-of-the-envelope calculations show that on the basis of our instrumental variables estimates, a doubling of the number of farmers markets in the average state-year would be associated with an economic cost of about $1.1 million in that state-year alone.

The remainder of this paper is organized as follows. Section 2 presents the data and discusses descriptive statistics. In section 3, we lay out our empirical framework, discuss our identification strategy, and outline the various robustness checks we conduct. Section 4 presents and discusses our estimation results and, perhaps more importantly, includes a discussion of the limitations of our findings and of how they should—and should not—be interpreted. We conclude in section 5 by discussing the policy implications of our findings and by providing directions for future research.

2. Data and Descriptive Statistics

The data we use come from administrative sources. The data on reported outbreaks and cases of food-borne illness are from the US Centers for Disease Control and Prevention's (CDC) Foodborne Outbreak Online Database (FOOD), which cover the years 1998 to 2013. All types of outbreaks included in the data are retained for analysis. We also include multistate outbreaks and related cases by ascribing them to the relevant states, but those were only available upon request from the CDC, and then again only for the years 2009-2013.[4] Because the presence of multistate outbreaks and related cases only for the period 2009-2013 complicates our analysis, we devote part of section 3 to how we deal with those multistate outbreaks and related cases.

Before we further discuss the data used in this paper, an important clarification needs to be made. At this point, it would be natural for the reader to ask whether it is possible to actually link an outbreak

[4] Though it might seem odd at first glance to talk of "multistate cases," we use that expression to refer to cases of food-borne illness that are associated with multistate outbreaks, and which are not included in the publicly available data.

or case of food-borne illness to the specific point of sale of the contaminated food. As it turns out, this is not feasible. The CDC's FOOD reports the place where the food that has caused an outbreak or a case of food-borne illness was *prepared*—at home, a restaurant, school, and so on—but not where it was sold. Moreover, the CDC data report the location of consumption in some but far from all cases. This means two things: First, it is not possible to ascribe an outbreak or a case of food-borne illness to a specific point of sale—for example, a farmers market. Second, this highlights the importance of the fact that foods consumed at home, at a restaurant, or at school can be purchased anywhere, which makes it even more difficult to ascertain a direct link between farmers markets and food-borne illness.

The data on farmers markets come from the USDA's Agricultural Marketing Service and include all farmers markets in the USDA's Farmers Markets Directory for the years 2004, 2006, and 2008-2013. Because data on farmers markets were not available for the years 2005 and 2007 or before 2004, the overlap between the food-borne illness data and the farmers markets data covers all 50 states and the District of Columbia for eight years—2004, 2006, and 2008-2013—for a total sample size of 408 observations.

The number of food-borne illness outbreaks in a given state in a given year is almost surely underreported. For an outbreak to be recorded in the CDC's FOOD, it has to be reported to the CDC by the relevant county authorities, who rely on medical personnel reports, who in turn rely on people deciding to go to medical facilities for treatment when they exhibit certain symptoms. Often, however, people suffering from food-borne illness might not consult medical personnel for their illness, and the medical personnel they interact with might not report that illness to county authorities, who may or may not report it to the CDC. The almost certain systematic underreporting of the dependent variable is discussed below, when we discuss our identification strategy, as a possible source of bias.

When unbundling food-borne illness by type, we initially chose to retain the top 10 food-borne illnesses in terms of outbreaks in the data. Because the top 10 includes four different varieties of norovirus, however, we chose to aggregate all those types of norovirus into one "all-norovirus" category. Starting from the top 10 food-borne illnesses, we thus end up with the top seven food-borne illnesses in our data. In order of importance, those are norovirus, *Salmonella enterica*, *E. coli shiga*, *C. perfringens*, *Campylobacter jejuni*, the scombroid toxin, and *Staphylococcus aureus*.

Regarding our control variables, state gross domestic product (GDP) figures are from the US Bureau of Economic Analysis. State population figures for 2004, 2006, and 2008-2009 are from the US Census Bureau's Population Division, but the figures for 2010-2013 are from the Census Bureau's American Community Survey (ACS). Likewise, college graduation rates are from the US Census Bureau's Current Population Survey (CPS) for 2004, but from the ACS for 2006 and 2008-2013, since the CPS data are not available for 2004. The data on the number of restaurants per state are from the US Census Bureau.[5] We include proxies for education and income given that those have been found to explain the location of farmers markets (Berning et al., 2013), and we include the number of restaurants per million given that those account for the amount of food consumed away from home. Finally, when it comes to the data we use for our placebo test, the number of bankruptcy filings per state are from the American Bankruptcy Institute.

Lastly, the variable we use as a source of plausibly exogenous variation for the number of farmers markets per million in a given state-year measures the average minimum daily temperature in a given state-year. We tried a few other weather-related variables—degree days above 0, 10, and 30 degrees Celsius, average maximum daily temperature, and even rainfall—but only average minimum daily temperature was correlated highly enough with (that is, a strong enough instrument for) the number of

[5] For those variables coming from more than one source, our use of year fixed effects obviates concerns about comparability between sources.

farmers markets per million. These are the same data used Tack et al. (2012) and by Schlenker and Roberts (2009) to study the relationship between climate change and crop yields in the US. The shortcoming of the weather data is that it does not cover the states of Alaska and Hawaii or the District of Columbia. This means that we lose 24 observations when incorporating weather as an instrument (i.e., our sample size goes from n=408 to n=384).

Table 1 presents descriptive statistics for all 50 states and the District of Columbia for the years 2004, 2006, and 2008-2013 for the dependent variables (i.e., all reported outbreaks and cases of food-borne illness as well as the reported number of reported outbreaks and cases for the seven most common illnesses), for the variable of interest (i.e., the number of farmers markets), and for the control variables.

The average state-year reports 21 outbreaks and 354 cases of food-borne illness per year, which includes about six outbreaks and 158 cases of norovirus, four outbreaks and 52 cases of *Salmonella enterica*, one outbreak and seven cases of *E. coli shiga*, and fewer than one outbreak of *Clostridium perfringens* (27 cases), *Campylobacter jejuni* (nine cases), scombroid (one case), and *Staphylococcus aureus* (five cases). The average state-year also has a total of 116 farmers markets. The average state-year has a GDP of $289 billion, a little over one fourth of its population has a college degree, it has a little over 11,000 restaurants, and a population of almost 6 million. Finally, the average daily minimum temperature in the average state-year for the period we study was equal to 5.46 degrees Celsius, or about 42 degrees Fahrenheit.

3. Empirical Framework

We begin this section by discussing our equation of interest, whose estimation relies on standard linear methods. We then discuss our identification strategy, which first exploits the longitudinal nature of our data, and then exploits weather shocks as a source of plausibly exogenous variation to explain the number of farmers markets per capita in a given state-year. Finally, we discuss the additional estimation strategies, both parametric and semiparametric, we rely on in order to ensure that our findings are robust.

3.1. Estimation Strategy

The equation of interest is such that

$$y_{it} = \alpha + \beta x_{it} + \gamma D_{it} + \delta_i + \tau_t + \epsilon_{it}, \tag{1}$$

where y is one of our outcomes of interest (i.e., the number of reported food-borne illness outbreaks or cases of food-borne illness, either total of by type of illness, per million in state i in year t), x is a vector of control variables, D is the treatment variable (i.e., the number of farmers markets per million in state i in year t), δ is a vector of state fixed effects, which control for all the time-invariant factors within each state, τ is a vector of year fixed effects, which control for all the state-invariant factors within each year, and ϵ is an error term with mean zero.

Our goal is to estimate γ which, if D were randomly assigned, would measure the causal effect of increasing the number of farmers markets per million by one on the reported number of outbreaks or cases of food-borne illness per million in the average state-year. Consequently, our statistical test of interest consists in testing the null hypothesis $H_0: \gamma = 0$ versus the alternative hypothesis $H_A: \gamma \neq 0$.

One immediate complication stems from the fact that our dependent variables are measured differently in different years. That is, for the years 2004, 2006, and 2008, the reported numbers of multistate outbreaks and related cases exclude multistate outbreaks and related cases. To remedy this, we estimate two additional specifications of equation (1). The first specification assumes that the inclusion of multistate outbreaks and related cases only affects the intercept of equation (1), such that we estimate

$$y_{it} = \alpha + \beta x_{it} + \gamma D_{it} + \theta m_t + \delta_i + \tau_t + \epsilon_{it}, \tag{1'}$$

where m_t is a dummy variable equal to one if multistate outbreak data were not recorded (i.e., missing) in year *t* and equal to zero if multistate outbreak data were recorded in year *t*. The second specification assumes that the inclusion of multistate outbreaks and related cases affects both the intercept as well as the slope of equation (1), such that we estimate

$$y_{it} = \alpha + \beta x_{it} + \gamma_D D_{it} + \gamma_{Dm}(D_{it} \cdot m_t) + \theta m_t + \delta_i + \tau_t + \epsilon_{it}, \tag{1''}$$

where m is defined as in equation (1’). For the specification in equation (1’’), we report an estimate of $\frac{\partial y}{\partial D} = \gamma_D + \gamma_{Dm}\bar{m}$ whenever applicable (i.e., the marginal effect of the number of farmers markets per million on the relevant dependent variable), where $\bar{m}$ is the sample mean of m_t (i.e., 0.375, given that multistate outbreak data are missing for 2004, 2006, and 2008, or three out of eight years in our sample). In what follows, we use the specification in equation (1’) as our core results, with the specifications in equations (1) and (1’’) shown for robustness.

Given just how unlikely it is that there exists a relationship between farmers markets and food-borne illness a priori at such an aggregated level as the state level, a rejection of the null in either direction should constitute *prima facie* evidence in favor of the hypothesis that there might be a relationship between farmers markets and food-borne illness. Yet several issues arise that compromise the identification of such a relationship. The next section discusses these issues, and the strategy we rely on in our effort to disentangle causation from correlation.

3.2. Identification Strategy

Many factors compromise the identification of γ in equation 1. Those factors can be grouped under three broad sources of statistical endogeneity, viz. (i) reverse causality, (ii) unobserved heterogeneity, and (iii) measurement error. In what follows, we first discuss our primary identification strategy, and we then discuss each of those potential sources of bias in turn. We then discuss the instrumental variable we use in order to tease out the potential causal relationship flowing from farmers markets to food-borne illness.

Recall that equation 1 includes a vector δ of state fixed effects, which control for all the time-invariant factors in a given state, and a vector τ of year fixed effects, which control for all the state-invariant factors in a given year.

State fixed effects in equation 1 allow purging the error term of its endogeneity due to unobserved, time-invariant heterogeneity across states. Similarly, year fixed effects allow purging the error term of its prospective endogeneity due to unobserved, state-invariant heterogeneity across years. Thus, to bias our estimate of γ, any remaining heterogeneity must either (i) vary systematically over time across states, or (ii) vary systematically across states over time and (iii) not be accounted for by the variables on the right-hand side (RHS) of equation (1).

We now turn to the sources of statistical endogeneity that compromise the identification of γ and discuss them in turn. In the case of reverse causality, though there is little doubt that y and D are jointly determined (i.e., they are likely to be both affected by a common set of unobserved confounders), it is possible that the number of farmers markets per million in a given state-year is caused by the number of reported outbreaks of food-borne illness per million in the same state-year. For one, Bond et al. (2006, 2008, and 2009), Smithers et al. (2008), and Thilmany et al. (2008) note that consumers often shop at farmers markets because they believe that the foods they purchase there are safer. Intuitively, however, those results would make it more likely that more consumers would shop at existing farmers markets because of safety concerns (i.e., the intensive margin within each farmers market) than it would drive the demand for farmers markets enough so as to increase the actual number of farmers markets (i.e., the extensive margin).

But perhaps more importantly, the rise in popularity of farmers markets over the last 10 to 20 years was due almost entirely to a growing taste for freshness, to a preference for spending money locally, and to the belief that local foods are more healthful when it comes to long-term health (whether that belief is right or wrong) rather than because of a belief that local foods are less likely lead to outbreaks of food-borne illness. Indeed, in their study of food hygiene and safety at farmers markets, Worsfold et al. (2004) found that consumers mainly care about product quality, and show little to no concern about food safety, and Toler et al. (2009) find evidence that other-regarding preferences and a preference for fairness can explain part of the rise of farmers markets. Similarly, Lusk (2015) finds that almost 45 percent of respondents have no opinion as to whether the foods from farmers markets increase or decrease the risk of food-borne illness, about 27 percent of respondents believe that the foods from farmers markets are more risky than other foods, and about 27 percent of respondents believe the contrary. In other words, according to Lusk's (2013) nationally representative data, respondent beliefs in one direction appear to cancel respondent beliefs in the other direction, which means that the overall effect of reverse causality is, in principle, nil. Given that Lusk's sample is representative of the US population, it appears unlikely that reverse causality from food-borne illness to farmers markets would bias our results one way or the other, especially given that we control for state fixed effects, year fixed effects, and other state-year-specific controls.

In the case of unobserved heterogeneity, the combined use of state fixed effects and year fixed effects should eliminate most unobserved heterogeneity between state-year observations. Indeed, state fixed effects purge the error term of its correlation with the treatment variable due to things that remain constant over the period 2004, 2006, and 2008-2013 for a given state (e.g., each individual state's proclivity to have fewer or more farmers markets), and year fixed effects purge the error term of its correlation with the treatment variable due to things that remain constant across all states in a given year (e.g., a legislative change that makes farmers markets more easily established across the country, or a country-wide outbreak of a specific food-borne illness). The identifying assumption we make here is thus that whatever unobserved heterogeneity is left does not significantly bias our estimate of γ. But because we cannot completely rule out the possibility that there is unobserved heterogeneity in our data which varies systematically across states and over time, we go a step further by also estimating specifications that include (i) a linear time trend, (ii) state specific-trends, (iii) US Census Bureau regional division-specific linear time trends, and (iv) US Census Bureau regional division-year fixed effects, all in an effort to eliminate the potential bias stemming from such unaccounted-for unobserved heterogeneity.

Lastly, in the case of measurement error, recall that the number of food-borne illness outbreaks and cases in each state-year is almost surely underreported, for the reasons discussed above. Although systematic measurement error is a usually threat to the identification of γ, note that in this case, this would mean that our estimate $\hat{\gamma}$ of γ would be such that $|\hat{\gamma}| < |\gamma|$. Indeed, if the number of outbreaks or cases of food-borne illness is underreported, the estimated relationship between the number of farmers markets and outbreaks would suffer from attenuation bias (i.e., it would be biased toward zero), because it would fail to account for a number of missing instances of the outcome variable. In other words, the measurement error just described would make one less likely to reject the null hypothesis $H_0\colon \gamma = 0$, which means that a rejection of the null in either direction would constitute a stronger result in this context, given that $\hat{\gamma}$ is an estimate of the lower bound on the true effect γ.

To recapitulate, reverse causality should not be a source of bias in this context, unobserved heterogeneity is largely controlled for by the combined use of state and year fixed effects (as well as trends and other kinds of fixed effects), and although there is measurement error in the dependent variable, that measurement error would tend to bias the estimate of the average treatment effect in equation (1) toward zero, which means that a rejection of the null hypothesis that $\gamma = 0$ makes for a stronger statement regarding the true relationship and that the estimated γ is an estimate of the lower bound on the true effect of farmers markets on food safety.

There remains one last threat to identification, viz. a violation of the stable unit treatment value assumption (SUTVA; cf. Pearl, 2009). In this context, the SUTVA states that the number of farmers markets in a given state-year should have no impact on the number of food-borne illness outbreaks in another state-year. To partially control for violations of the SUTVA, recall that we account for multistate outbreaks and related cases in our data. Moreover, we estimate specifications wherein we control for the number of farmers markets per million in neighboring states. We do this in order to control for within-year spillovers. For example, residents of western Wisconsin often shop in the Twin Cities of Minneapolis-Saint Paul given the relatively short distance between the two, and it is not completely unlikely that foods purchased in Saint Paul, MN might cause a case or outbreak of food-borne illness in, say, Hudson, WI. The foregoing only controls for SUTVA violations that might occur between states within a given year, and not for SUTVA violations that might occur within a state over time, or between states over time. On the former, given the relatively short-lived nature of episodes of food-borne illness, we believe it is unlikely that a late December shopping trip to the farmers market could lead to a case of food-borne illness reported in January in the data, if only because in most of the United States, few to no farmers markets operate that time of year. Even if there were such cases, it is unlikely that there would be enough of them that they would bias our estimates. On the latter, this is a clear shortcoming of our analysis, but given the relatively small sample size of 408 state-year observations, we think it best not to torture the data by imposing the structure required to model interstate dynamics, which in any event would require strong functional form assumptions that we are not willing to make.

Moreover, in order to disentangle a potentially causal relationship between farmers markets and food-borne illness from the correlation between the two, we estimate a two-stage least squares (2SLS) version of equation (1’) in which we instrument the number of farmers markets per million in a state-year with the average minimum daily temperature in the same state-year.

Our rationale for using the average maximum daily temperature as an instrumental variable (IV) is as follows. First off, on the exogeneity front, by incorporating state fixed effects and a linear time trend,[6] we are identifying off of shocks to average minimum daily temperature within each state taking into account the passage of time. Because they measure deviations from the within-state average daily minimum temperature via state fixed effects and linear trends account for the possible change over time in those deviations, our IV is unpredictable, and thus plausibly exogenous to the number of farmers markets in a given state-year, and thus uncorrelated with the error term in equation (1’).

Second, on the relevance front, shocks to average daily minimum temperature should influence the within-state mean in the number of farmers markets per capita in two ways. It should do so directly because negative and positive shocks to average minimum daily temperature make it respectively less and more likely that new farmers markets will open in a given state. It should also do so indirectly, because those same shocks to average minimum daily temperature respectively decrease and increase the yields of crops typically sold at farmers markets (e.g., fruits and vegetables), which might at the margin affect whether fewer or more farmers markets open in a given year. Because this latter channel might violate the exclusion restriction, and thus make our IV only plausibly rather than strictly exogenous—as crop yields change in response to weather shocks, so do those crops’ availability and prices, which might affect what people buy and eat, and thus their exposure to different types of food-

[6] Though we did try to estimate 2SLS specifications wherein we control for both state and year fixed effects, the inclusion of year fixed effects makes our IV so weak as to be useless given that shocks to average minimum daily temperatures tend to be correlated across states within a given year. As such, the best we can do for our 2SLS specifications is to control for state fixed effects and a linear trend. Moreover, this means that by virtue of controlling for the variation provided for by our instrument, the estimates from our OLS specifications are likely close to the average treatment effect.

borne illness—we assess the robustness of our IV estimates using the method laid out by Conley et al. (2012).

Given the foregoing, the relationship between farmers markets and the IV is likely to be monotonic. This means that our 2SLS specifications identify local average treatment effects—that is, the effect of farmers markets on food-borne illness *in those states where negative and positive weather shocks induced fewer or more farmers markets to open*. As such, the magnitude of the LATEs are not directly comparable with the magnitude of the ATEs obtained from our OLS regressions.

Before concluding this section, we wish to impress upon the reader the difficulty posed by finding a valid instrument for farmers markets in this context. Indeed, whether a farmers market opens at all in a given state in a given year is often determined by factors determined in the same state, but in previous years. Properly exploiting this would require the use of lagged variables as controls or as IVs, which Bellemare et al. (2017) show often does more to compromise the identification of causal effects (and lead to wrong inference) than make it possible. Focusing on contemporaneous (i.e., non-lagged) variables, we turned to weather-based potential IVs. After also considering the average maximum daily temperature, average precipitation, and degree days above 0, 10, or 30 degrees Celsius in each state-year as IVs, but none of those variables were strong instruments, average minimum daily temperature was the only variable that had enough explanatory power in the first-stage equation, i.e., it was the only instrument that was not weak.

Finally, in line with Bertrand et al.'s (2004) conclusions and the recommendations in Angrist and Pischke (2009, 2014) and Cameron and Miller (2015), we cluster our standard errors at the state level throughout to make our results robust to general forms of heteroskedasticity and autocorrelation. We also conduct placebo and falsification tests by respectively (i) regressing each of our outcome variables on a fake treatment (i.e., the number of bankruptcies per million in a given state in a given year), and (ii) regressing the number of bankruptcies per million in a given state on the RHS variables in equation (1').

3.3. Alternative Estimators

After presenting the results of the linear model in equation 1 as well as a battery of robustness checks, we estimate a nonlinear variant of our core equation. This alternative version of our core equation consists of a semiparametric specification wherein we allow for potential nonlinearities in the relationship between the number of farmers markets and each of the reported number of (i) outbreaks of food-borne illness, (ii) cases of food-borne illness, (iii) outbreaks of norovirus, (iv) cases of norovirus, (v) outbreaks of Campylobacter jejuni, and (vi) cases of *Campylobacter jejuni*. Specifically, we estimate a spline regression (Härdle, 1990; Yatchew, 2003), which entails estimating the following, modified version of equation 1:

$$y_{it} = \alpha + \beta x_{it} + \gamma f(D_{it}) + \delta_i + \tau_t + \epsilon_{it}. \tag{2}$$

All variables in equation (2) are the same as in equation (1'), the only difference being that our treatment variable D now enters the RHS through a nonlinear function $f(\cdot)$. The function $f(\cdot)$ we choose to estimate here is a spline with five knots, which affords a greater amount of flexibility than specifications with fewer (e.g., three) knots while minimizing the curse of dimensionality associated with specifications with more (e.g., seven) knots. Given that this is a nonlinear procedure, rather than presenting a table of results, we present for each semiparametric regression a figure showing the estimated nonlinear relationship between the number of farmers markets per million and the extent of food-borne illness per million.

4. Estimation Results and Discussion

We begin this section by presenting and discussing scatter plots showing unconditional relationships between farmers markets and food-borne illness. Because those scatter plots fail to account for confounding factors, the bulk of this section focuses on our core results, viz. our results for the linear, parametric specifications in equation (1’) and all relevant robustness checks. We then discuss estimation results for our placebo and falsification tests before moving on to the tobits and the spline regressions discussed in the previous section. We conclude this section with a discussion of the limitations of our findings.

4.1. Scatter Plots

Figures 1a and 1b respectively show scatter plots wherein the number of farmers markets per million is shown on the *X*-axis, the number of outbreaks of food-borne illness per million (in Figure 1a) and the number of cases of food-borne illness per million (in Figure 1b) is shown on the *Y*-axis, and each point in the scatter represents one state in a given year over the period 2004, 2006, and 2008-2013. In both figures, a linear regression of the dependent variable on the number of farmers markets per million is included, along with its 95 percent confidence interval. In both cases, there is a positive unconditional relationship between farmers markets and food-borne illness.

The scatter plots in Figures 1a and 1b obviously mask a great deal of heterogeneity, which we tackle more fully in the next sub-section. But before doing so, Figures 2a to 2h show similar scatter plots to that in Figure 1a. Namely, Figures 2a to 2h plot the number of farmers markets per million on the *X*-axis and the number of reported outbreaks of food-borne illness per million on the *Y*-axis for each state in a single year. Figures 2a to 2h thus show that the positive relationship between the number of farmers markets per million and the number of reported outbreaks of food-borne illness per million identified in Figure 1a not only holds over the entire period 2004, 2006, and 2008-2013—it also holds for each and every one of the eight years covered by our data. Similar annual plots for the number of reported cases of food-borne illness per million (not shown, but available upon request) yield similar results.

4.2. Linear Regressions

Tables 2 and 3 present estimation results for equation (1’) for eight outcomes of interest. Table 2 presents estimation results for reported outbreaks of food-borne illness; Table 3 presents estimation results for reported cases of food-borne illness. The first column of each table reports results for the total reported number of outbreaks or cases of food-borne illness; the remaining seven columns report results for the reported number of outbreaks or cases of specific illnesses.

Tables 2 and 3 tell a similar story: When controlling for state fixed effects as well as year fixed effects in addition to the control variables discussed in section 2, there is a positive, significant relationship between the number of farmers markets per million on the one hand and the total reported number of outbreaks or cases of food-borne illness per million on the other hand. Moreover, there are similar such relationships for norovirus and *Campylobacter jejuni*.

From the results in the first column of Table 3, we can also take a first pass at a back-of-the-envelope calculation aimed at assessing the cost of food-borne illness associated with farmers markets. The marginal effect in this case is equal to 1.164. With an average of 29.67 farmers markets per million in the average state-year, this means that a doubling of farmers markets—and recall that the number of farmers markets in the US has more than quadrupled between 1994 and 2014—would lead to 34.54 (i.e., 29.67 × 1.164) additional cases of food-borne illness per million. Using Scharff's (2012) $1,068-per-case-of-food-borne-illness economic cost figure, this would mean that a doubling of the number of farmers markets would imply an economic cost of $36,889 per million. And with a population of 5.95 million individuals in the average state-year, this means that a doubling of the number of farmers markets would be associated with an economic cost of about $220,000 in the average state-year. Unfortunately, we cannot recover a similar measure of cost per reported outbreak or of cost per reported outbreak or case of norovirus or *Campylobacter jejuni*, since estimates of the economic costs of those events are unavailable.

Appendix Tables A1 to A4 show results for both equations (1) and (1''). For the results for equation (1'') in Tables A2 and A4, there are two coefficients, i.e., γ_D and γ_{Dm}, which we combine into an estimate $\widehat{\partial Y/\partial D} = \hat{\gamma}_D + \hat{\gamma}_{Dm}\overline{m}$ of the marginal effect of the number of farmers markets per million in the average state-year on the relevant dependent variable. This marginal effect is shown, clearly highlighted, at the bottom of tables A2 and A4. Our core results (i.e., the results in tables 2 and 3) are robust to those different specifications.

Given the lack of statistical significance between the number of farmers markets per million and cases or outbreaks of *Salmonella enterica*, *E. coli shiga*, *C. perfringens*, scombroid toxin, and *Staphylococcus aureus* per million in Tables 2 and 3, we ignore those "dogs that did not bark" in the remainder of this paper.

Tables 4 to 9 assess the robustness of the core results shown in Tables 2 and 3 by looking at specifications wherein we (i) control for the number of farmers markets in neighboring states divided by the population in the relevant state-year to control for spillovers,[7] (ii) control for a linear trend instead of year fixed effects, (iii) control for state-specific trends instead of year fixed effects, (iv) incorporate Census Bureau regional division-year fixed effects on top of state fixed effects, and (v) we revert to the specification in (ii) but instrument the number of farmers markets using the average minimum daily temperature.

Thus, the specifications in columns (1) to (4) present OLS estimates, whereas the specification in column (5) of Tables 4 to 9 presents 2SLS estimates. As discussed in the previous section, the results of OLS and 2SLS specifications are not directly comparable, but looking at the OLS specifications in columns (1) to (4) of Tables 4, 6, and 8, the estimates of the relationship between farmers markets and outbreaks of food-borne illness appear remarkably stable. Following Altonji et al. (2005), this suggests that there is little omitted variables bias. The estimates of the relationship between farmers markets and cases of

[7] For example, to account for spillovers from other states for Minnesota in 2013, we count the number of reported outbreaks of food-borne illness in North Dakota, South Dakota, Iowa, and Wisconsin in 2013 and divide this total by the population of Minnesota in 2013.

food-borne illness in Tables 5 and 7 are less stable, perhaps owing to the considerably noisier nature of the cases data. Looking at the results in Table 9, the relationship between farmers markets and cases of campylobacter disappears. In sum, our core results are largely robust to the additional specifications in columns (1) to (4) for total reported outbreaks and cases of food-borne illness (Tables 4 and 5), for reported outbreaks and cases of norovirus (Tables 6 and 7), and for outbreaks of campylobacter (Table 8).

Turning to the 2SLS results in column (5) of Tables 4 to 9, note that only the results for total reported outbreaks and cases of food-borne illness (Tables 4 and 5) and for reported outbreaks and cases of norovirus (Tables 6 and 7) are robust in their 2SLS specifications. Given a back-of-the-envelope analysis of the cost of food-borne illness seemingly caused by farmers markets similar to the one above, it looks as though a doubling of the number of farmers markets in the average state year would cause an economic cost of \$1.1 million. Specifically, this doubling would lead to 173.1 (i.e., 29.67×5.834) additional cases of food-borne illness per million. Using Scharff's (2012) \$1,068-per-case-of-food-borne-illness economic cost figure once again, this would mean that a doubling of the number of farmers markets would imply an economic cost of \$184,865 per million. Multiplying this figure by 5.95 million individuals in the average state-year yields the \$1.1 million figure. Again, we stress that this is a LATE. Thus, this \$1.1 million figure only applies to those state-years where the number of farmers markets increases or decreases in response to positive or negative shocks to the average daily minimum temperature (we have no way of determining the observations for which this is true given our data), whereas the \$220,000 figure above applies to our entire sample for the period 2004-2013. There is thus a trade-off between the internal and external validity of those cost figures.

In order to assess the robustness of our 2SLS results to small departures from the strict exogeneity assumption required for those results to be identified, we apply the method developed by Conley et al.

(2012) to deal with plausibly—but not strictly—exogenous instruments.[8] In applying this methodology, it is necessary to impose some sort of prior on said departures from strict exogeneity, with the tradeoff being that the least structured that prior, the less precise the estimates returned by the methodology. We thus pick Conley et al.'s intermediate local-to-zero (LTZ) method, which only requires that one impose a prior on the mean and standard deviation for their γ parameter (as distinct from ours above), which measures the magnitude of the presumed departure from strict exogeneity. In this case, we assume a mean of zero and a standard deviation of one—that is, we assume that, in expectation, strict exogeneity holds, but we allow for relatively wide departures from strict exogeneity.

The results of the Conley et al. (2012) LTZ checks show that only the 2SLS results for the total reported number of cases of food-borne illness (column 5 of Table 5) and for the reported number of cases of norovirus (column 5 of Table 7) are robust to departures from the assumption of strict exogeneity of average daily minimum temperature to food-borne illness. The 2SLS estimate of the relationship between the number of farmers markets per million and the total reported number of cases of food-borne illness per million is found by the LTZ method to be in the $[.2261093, 11.14962]$ 95% confidence interval; those of the relationship between the number of farmers markets per million and the reported number of cases of norovirus is found to be in the $[1.352072, 12.11773]$ 95% confidence interval. It thus looks as though our 2SLS results for those two outcome variables—the total reported number of cases of food-borne illness, and the reported number of cases of norovirus—are robust to small departures from the strict exogeneity assumption.

In Table 10, we report the results of placebo tests wherein we inflict a fake treatment (here, the number of bankruptcies per million) on our dependent variables in columns 1 to 6, and of a falsification

[8] Nevo and Rosen (2012) also develop a method to deal with what they dub imperfect instrumental variables. We chose Conley et al.'s method over Nevo and Rosen's because the latter involves making an assumption we were not willing to make, viz. the assumption that the sign of the correlation between the treatment variable and the error term is the same as the sign of the correlation between the instrumental variable and the error term.

test wherein we regress the number of bankruptcies per million on the number of farmers markets per million and the control variables on the RHS of equation 1. The results of all six placebo tests, which show no statistical significance for the number of bankruptcies per million for any of the six dependent variables, suggest that the estimation results in Tables 2 to 9 are unlikely to be spurious. The lack of statistical significance for the number of farmers markets in column 7 of Table 10 yields a similar insight.

Figures 3a and 3b show estimation results for the semiparametric (i.e., spline) regressions discussed in section 3, respectively plotting the estimated nonlinear relationship between the number of farmers markets per million and all reported outbreaks and cases of food-borne illness per million, along with confidence intervals. (The splines were estimated on centered data, i.e., data that were demeaned using the within-state mean of each variable to account for state fixed effects without having to deal with 50 additional parameters. This explains why the range of the variables on the *X*- and *Y*-axes in Figures 3a and 3b take on both positive and negative values.) Appendix Figures A1 to A4 show similar results for outbreaks and cases of norovirus and *Campylobacter jejuni* per million. In all cases, there is a generally monotonically increasing relationship between the number of farmers markets per million and the dependent variable.

Tables 11 and 12 look at whether our core results for all reported outbreaks and cases of food-borne illness are robust to removing the furthest upward outlier (due to the left-censored nature of our dependent variables), the furthest leverage point (i.e., outliers for the treatment variable) to the right (given that there are almost no cases of state-year observations with no farmers markets), or both. Appendix Tables A5 to A8 show similar results for outbreaks and cases of norovirus and *Campylobacter jejuni*. In all cases, results are robust to removing outliers or leverage points, but not both. This latter result is not too concerning. Since our sample consists of *all* US states, it is harder to argue that any state-year observation should be taken out because it is an outlier in this case than, say, in the case where we would be considering a random sample of state-year observations.[9]

To summarize, our results indicate that

1. There is a robust, positive association between the number of farmers markets per million on the one hand and the number of (i) total reported outbreaks of food-borne illness, (ii) total reported cases of food-borne illness, (iii) reported outbreaks of norovirus, (iv) reported cases of norovirus, and (v) outbreaks of campylobacter per million.
2. When using average minimum daily temperature in a given state-year as an IV for the number of farmers markets in the same state-year, there appears to be a causal relationship flowing from the number of farmers markets per million to the number of (i) total reported outbreaks of food-borne illness, (ii) total reported cases of food-borne illness, (iii) reported outbreaks of norovirus, and (iv) reported cases of norovirus.
3. Entertaining the possibility that our IV might only be plausibly exogenous using Conley et al.'s (2012) local-to-zero methodology, only the presumably causal relationships flowing from the number of farmers markets per million to the number of (i) total cases of food-borne illness and (ii) reported cases of norovirus are robust to departures from strict exogeneity.
4. Depending on whether one considers the ATE or the LATE, the economic cost of food-borne illness associated with (in the case of the ATE) or seemingly caused by (in the case of the LATE) a doubling of farmers markets in a given state is respectively on the order of $220,000

[9] That being said, because Figures 2a to 2h show that Vermont is always a leverage point (i.e., it always has more farmers markets per million than other states), we also estimated specifications of the core equations in Tables 2 and 3 (not shown, but available upon request) which omitted Vermont; all results were robust to this omission, which indicates that our core are not driven by Vermont.

or \$1.1 million. Obviously, the LATE estimate is limited in its external validity, as it applies only to those states where weather shocks drove the number of farmers markets.[10]

4.3. Limitations

While the foregoing suggests that farmers markets play a role in causing food-borne illness, there are a number of ways in which our results could be misinterpreted by nonexperts. To prevent that, we clarify in this section just how one should interpret our results. This section thus discusses the limitations of our results.

First, even if one grants that (some of) the results reported in this paper might be causal, it would be a mistake to interpret those results as saying that the foods purchased at farmers markets are somehow more likely to make consumers ill than the foods purchased at grocery stores because of our results. This is because our results do not allow studying the precise causal mechanisms through which farmers markets may increase the number of cases and outbreaks of food-borne illness. Indeed, most food safety problems come from the mishandling of foods by consumers or by restaurant staff who prepare those foods for consumers (Paarlberg, 2013). As such, it is easy to imagine cases where consumers are more or less neglectful with foods purchased from farmers markets (e.g., by being less likely to wash produce from the farmers market, or by cooking eggs from the farmers market more thoroughly than eggs from the grocery store, and so on), which could explain our results. In other words, although the presence of farmers markets in a given state might well lead to more cases and outbreaks of food-borne illness, this paper cannot pinpoint the precise causal mechanisms through which this occurs.

Second, recall that our results included a number of what Sherlock Holmes would have referred to as “dogs that did not bark.” Indeed, there was no significant relationship between farmers markets and

[10] Our LATE estimates are further limited in their external validity given that the sample size decreases from n=408 to n=384 when going from OLS to 2SLS estimates, i.e., when going from ATE to LATE. Recall that because data on average daily minimum temperatures were not available for Alaska, the District of Columbia, and Hawaii, those states are not included in the 2SLS analyses.

outbreaks or cases of *Salmonella enterica*, *E. coli shiga*, *Clostridium perfringens*, scombroid toxin, or *Staphylococcus aureus*. For those illnesses, it is impossible to tell whether this represents evidence in favor of an absence of any relationship between farmers markets and cases or outbreaks, or whether this was due to an absence of evidence. As such, just as one should be cautious when interpreting our significant results, one should also be cautious when interpreting those null results.

Third, our estimates of the relationship between farmers markets and food-borne illness are all estimates of the lower bound on the true relationship given the measurement error issues discussed in sections 2 and 3. Given just how underreported the extent of food-borne illness is in the US, it is likely that the true effect is much larger than our reported estimates.

Fourth, our different estimates of the cost of food-borne illness associated with farmers markets apply to different cases. The $220,000 figure applies to all US states for the period 2004-2013; the $1.1 million figure applies only to the states wherein weather shocks drove the number of farmers markets for the same period (and then again, excluding Alaska, Hawaii, and the District of Columbia, given partial coverage of the weather data). Given the measurement error just discussed for our dependent variables, this also means that the economic cost figures in this paper are likely to be underestimated.

Fifth, our findings are limited in their external validity. Specifically, because we consider only the period 2004, 2006, and 2008-2013, our findings only apply to those years, and there is no guarantee that they would apply to years before 2004, to 2005 or 2007, or to years after 2013.

Finally, our analysis does not attempt to and cannot measure the benefits, in terms of health of otherwise, that might be attributable to an increase in farmers markets. Doing so would be well beyond the scope of this paper.

5. Summary and Conclusions

Using data on the number of food-borne illness outbreaks and cases and the number of farmers markets per million across the entire United States for the period 2004, 2006, and 2008-2013, we exploit variation over time and space to study the relationship between farmers markets and food-borne illness. Our results indicate that, once the unobserved heterogeneity between states and the unobserved heterogeneity between years are taken into account, there is a positive relationship between the number of farmers markets per million in a given state and the reported number of all outbreaks and cases of food-borne illness per million as well as the reported number of outbreaks and cases of norovirus and the number of outbreaks of *Campylobacter jejuni* in the same state. These results are robust to nonlinear specifications and to removing outliers or leverage points, and placebo and falsification tests indicate that they are unlikely to be spurious.

When instrumenting the number of farmers markets per million in a given state-year with a variable capturing weather shocks that appear to drive whether fewer or more farmers markets open in a given state in a given year, it appears that increases in the number of farmers markets cause increases in the total reported number of outbreaks and cases of food-borne illness and in the reported number of outbreaks and cases of norovirus. Allowing for departures from the assumption of strict exogeneity of our instrumental variable, only the results for the total reported number of cases of food-borne illness and for the reported number of cases of norovirus are robust.

Although the research design in this paper falls short of the gold standard of experimental evidence and the causal identification of the estimated relationships is threatened by a number of factors, the fact that it was *a priori* unlikely that there existed a statistically significant relationship between the treatment and outcome variables at such an aggregate level as the state level but that such a relationship was nevertheless found (and not only found to be robust, but also to be seemingly causal given 2SLS results for total reported cases of food-borne illness and reported cases of norovirus), combined with our falsification test, placebo tests, alternative specifications, and alternative estimators all enhance the credibility of our finding.

From a policy perspective, it would be a mistake to discourage people to purchase food from farmers markets on the basis of our results. Indeed, even if our estimated relationships between farmers markets and food-borne illness were causal beyond any reasonable doubt, we would not be able to determine the precise mechanisms through which those relationships operate. This points to another research direction for researchers interested in studying the relationship between farmers markets and food-borne illness, i.e., the mechanisms whereby farmers markets might cause food-borne illness. The study of those mechanisms will likely necessitate primary data collection at the level of the farmers markets themselves. In light of the findings in this paper, this strikes us as a worthy endeavor for future research.

References

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber (2005), “Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools,” *Journal of Political Economy* 113(1): 151-84.

Angrist, Joshua D., and Jörn-Steffen Pischke (2009), *Mostly Harmless Econometrics*, Princeton, NJ: Princeton University Press.

Angrist, Joshua D., and Jörn-Steffen Pischke (2014), *Mastering ‘Metrics*, Princeton, NJ: Princeton University Press.

Bellemare, Marc F., Takaaki Masaki, and Thomas B. Pepinsky (2017), “Lagged Explanatory Variables and the Estimation of Causal Effects,” *Journal of Politics* forthcoming.

Berning, Joshua P., Alessandro Bonanno, and Etemaadnia Hamideh (2013), “Farmers’ Markets Location Determinants: An Empirical Analysis in New England,” Working Paper, University of Georgia.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), “How Much Should We Trust Differences-in-Differences Estimates?,” *Quarterly Journal of Economics* 119(1): 249-275.

Bond, Craig A., Dawn Thilmany, and Jennifer K. Bond (2008), “Understanding Consumer Interest in Product and Process-Based Attributes for Fresh Produce,” *Agribusiness* 24(2): 231-252.

Bond, Jennifer K., Dawn Thilmany, and Craig A. Bond (2006), “Direct Marketing of Fresh Produce: Understanding Consumer Purchasing Decisions,” *Choices* 21(4): 229-235.

Bond, Jennifer K., Dawn Thilmany, and Craig A. Bond (2009), “What Influences Consumer Choice of Fresh Produce Purchase Location?,” *Journal of Agricultural and Applied Economics* 41(1): 61-74.

Cameron, A. Colin, and Douglas L. Miller (2015), "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources* 50(2): 317-372.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi (2012), "Plausibly Exogenous," *Review of Economics and Statistics* 94(1): 260-272.

Francis, Gillian A., Christopher Thomas, and David O'Beirne, (1999) "The Microbiological Safety of Minimally Processed Vegetables," *International Journal of Food Science & Technology* 34(1): 1-22.

Greene, William (2002), "The Bias of the Fixed Effects Estimator in Nonlinear Models," Working Paper, New York University.

Härdle, Wolfgang (1990), *Applied Nonparametric Regression*, Cambridge: Cambridge University Press.

Harrison, Judy A., Julia W. Gaskin, Mark A. Harrison, Jennifer L. Cannon, Renee R. Boyer, and Geoffrey W. Zehnder (2013), "Survey of Food Safety Practices on Small to Medium-Sized Farms and in Farmers Markets," Journal of Food Protection 76(11): 1989-1993.

Heckman, James J. (1981), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in Manski, Charles F. and Daniel McFadden, Eds., *Structural Analysis of Discrete Data and Econometric Applications*, MIT Press, Cambridge, MA.

Lusk, Jayson (2013), "Food Demand Survey (FooDS)–Technical Information on Survey Questions and Methods," Working Paper, Oklahoma State University.

Lusk, Jayson (2015), "Food Demand Survey (FooDS) – March 2015," http://jaysonlusk.com/blog/2015/3/12/food-demand-survey-foods-march-2015 last accessed July 16, 2015.

Lusk, Jayson (2016), *FooDS – Food Demand Survey* 4(2): 1-6.

Maddala, G.S. (1983), Limited Dependent and Qualitative Variables in Econometrics, Cambridge: Cambridge University Press.

Nevo, Aviv, and Adam M. Rosen (2012), “Identification with Imperfect Instruments,” *Review of Economics and Statistics* 94(3): 659-671.

Paarlberg, Robert (2013), *Food Politics*, 2nd edition, Oxford: Oxford University Press.

Pearl, Judea (2009), *Causality*, 2nd ed., Cambridge University Press.

Park, C.E., and G.W. Sanders (1992), “Occurrence of Thermotolerant Campylobacters in Fresh Vegetables Sold at Farmers' Outdoor Markets and Supermarkets,” *Canadian Journal of Microbiology* 38(4): 313-316.

Scharff, Robert L. (2012), “Economic Burden from Health Losses Due to Foodborne Illness in the United States,” *Journal of Food Protection* 75(1): 123-131.

Scheinberg, Joshua, Stephanie Doores, and Catherine N. Cutter (2013), “A Microbiological Comparison of Poultry Products Obtained from Farmers' Markets and Supermarkets in Pennsylvania,” *Journal of Food Protection* 33(3): 259-264.

Schlenker Wolfram, and Michael J. Roberts (2009), “Nonlinear Temperature Effects Indicate Severe Damages to US Crop Yields under Climate Change,” *Proceedings of the National Academy of Sciences* 106: 15594-15598.

Sivapalasingam, Sumathi, Cindy R. Friedman, Linda Cohen, and Robert V. Tauxe (2004), “Fresh Produce: A Growing Cause of Outbreak of Foodborne Illness in the United States, 1973 through 1997,” *Journal of Food Protection* 10: 2092-2353.

Smithers, John, Jeremy Lamarche, and Alun E. Joseph (2008), "Unpacking the Terms of Engagement with Local Food at the Farmers' Market: Insights from Ontario," *Journal of Rural Studies* 27: 337-350.

Tack, Jesse B., Ardian Harri, and Keith Coble (2012), "More than Mean Effects: Modeling the Effect of Climate on the Higher Order Moments of Crop Yields," *American Journal of Agricultural Economics* 94(5): 1037-1054.

Thilmany, Dawn, Craig A. Bond, and Jennifer K. Bond (2008), "Going Local: Exploring Consumer Behavior and Motivations for Direct Food Purchases," *American Journal of Agricultural Economics* 90(5): 1303-1309.

Toler, Stephen, Brian C. Briggeman, Jayson L. Lusk, and Damian C. Adams (2009), "Fairness, Farmers Markets, and Local Production," *American Journal of Agricultural Economics* 91(5): 1272-1278.

US Centers for Disease Control and Prevention (2014), "CDC Estimates of Food-Borne Illness in the United States," http://www.cdc.gov/foodborneburden/questions-and-answers.html last accessed June 17, 2014.

US Centers for Disease Control and Prevention (2016a), "Norovirus," http://www.cdc.gov/norovirus/ last accessed February 24, 2016.

US Centers for Disease Control and Prevention (2016b), "*Campylobacter*," http://www.cdc.gov/foodsafety/diseases/campylobacter/index.html last accessed February 24, 2016.

Worsfold, Denise, Philip M. Worsfold, and Christopher J. Griffith (2004), "An Assessment of Food Hygiene and Safety at Farmers' Markets," *International Journal of Environmental Health Research* 14(2): 109-119.

United States Department of Agriculture Agricultural Marketing Service (2015), "Farmers Markets and Local Food Marketing," http://www.ams.usda.gov/AMSv1.0/ams.fetchTemplateData.do?template=TemplateS&leftNav=WholesaleandFarmersMarkets&page=WFMFarmersMarketGrowth&description=Farmers+Market+Growth last accessed June 10, 2015.

Yatchew, Adonis (2003), *Semiparametric Regression for the Applied Econometrician*, Cambridge: Cambridge University Press

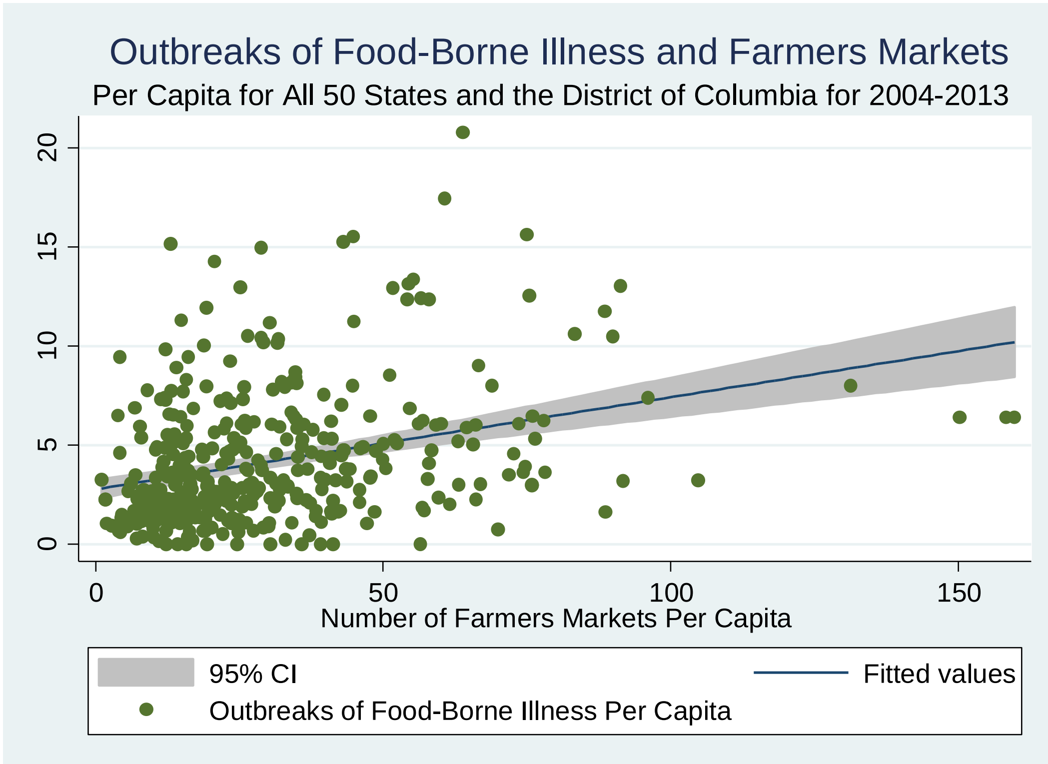


Figure 1a. Reported Outbreaks of Food-Borne Illness and Farmers Markets Per million, 2004-2013.

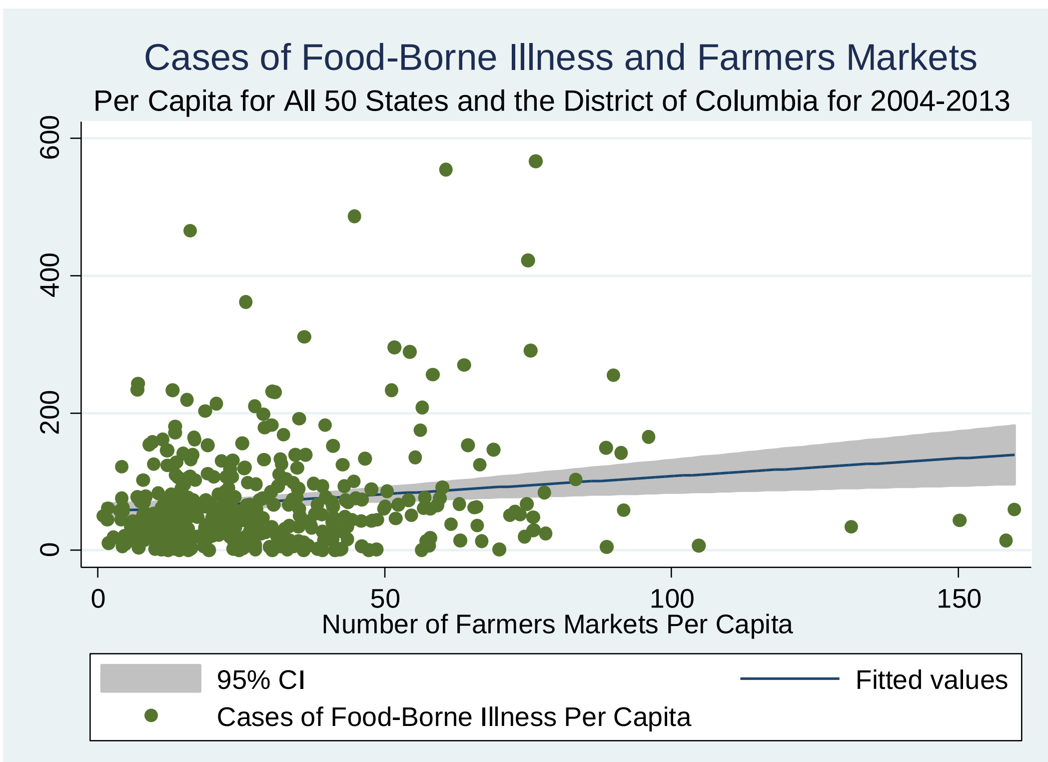


Figure 1b. Reported Cases of Food-Borne Illness and Farmers Markets Per million, 2004-2013.

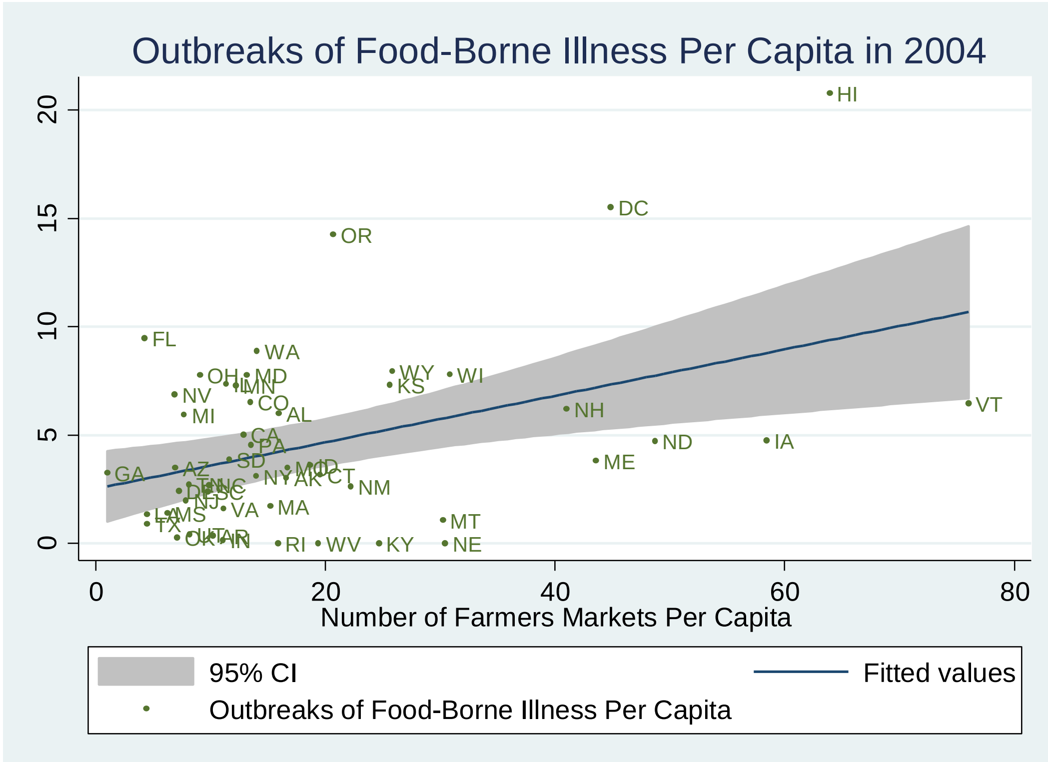


Figure 2a. Reported Outbreaks of Food-Borne Illness and Farmers Markets Per million in 2004.

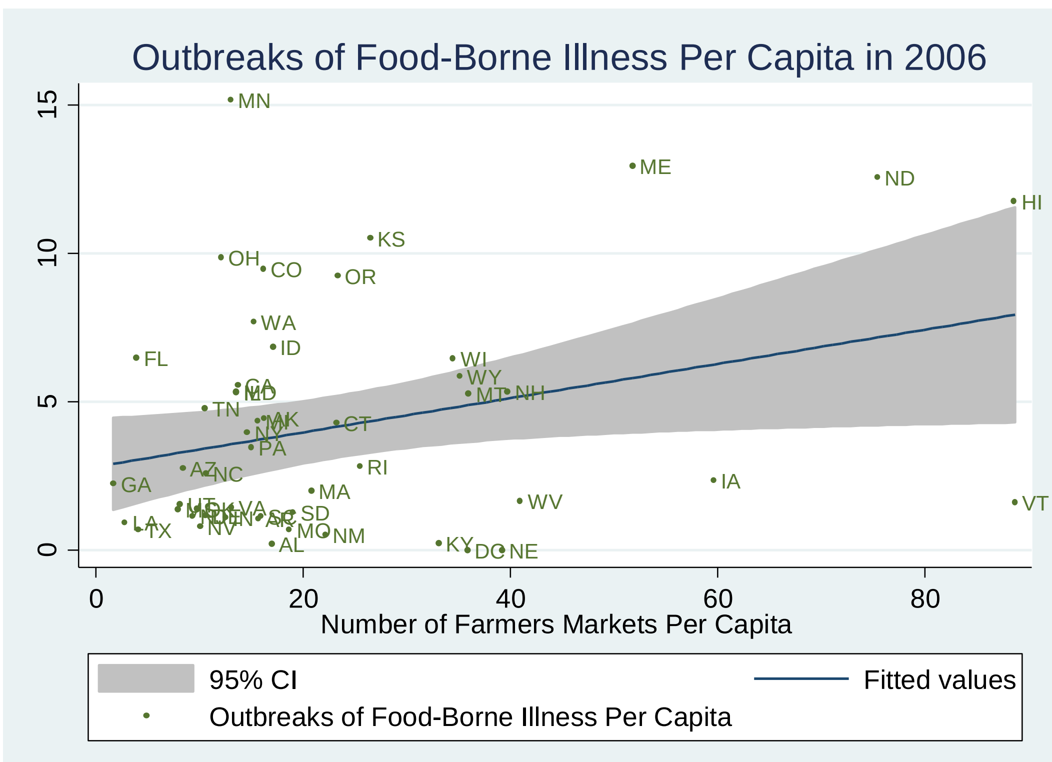


Figure 2b. Reported Outbreaks of Food-Borne Illness and Farmers Markets Per million in 2006.

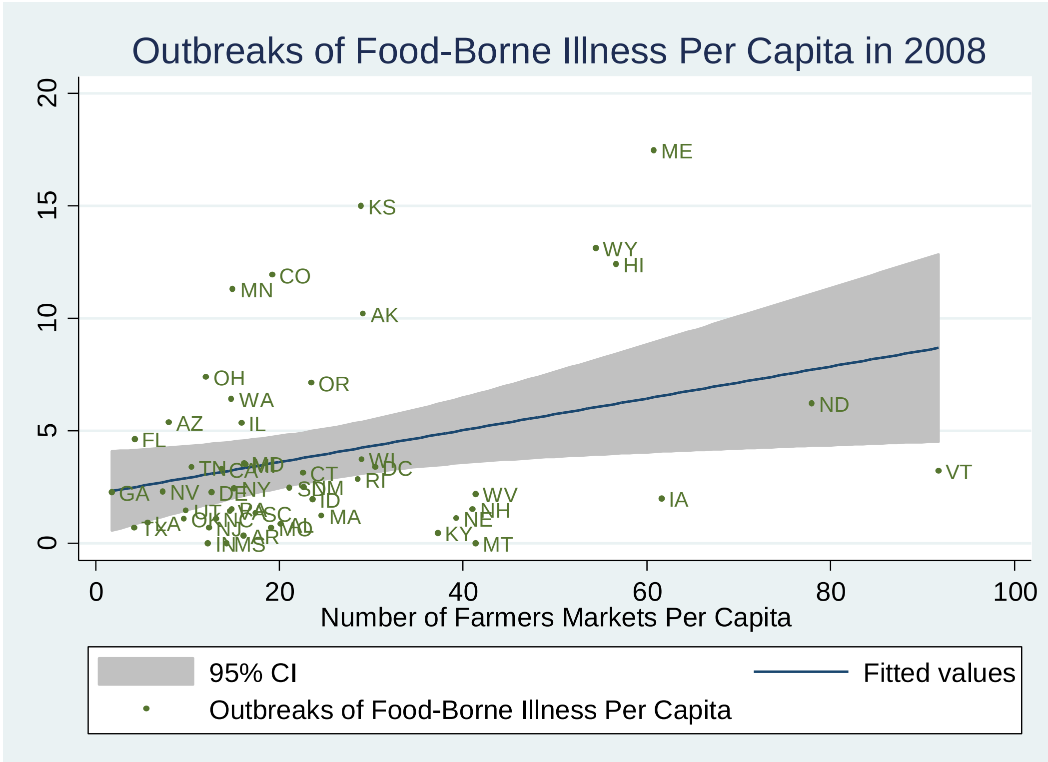


Figure 2c. Reported Outbreaks of Food-Borne Illness and Farmers Markets Per million in 2008.

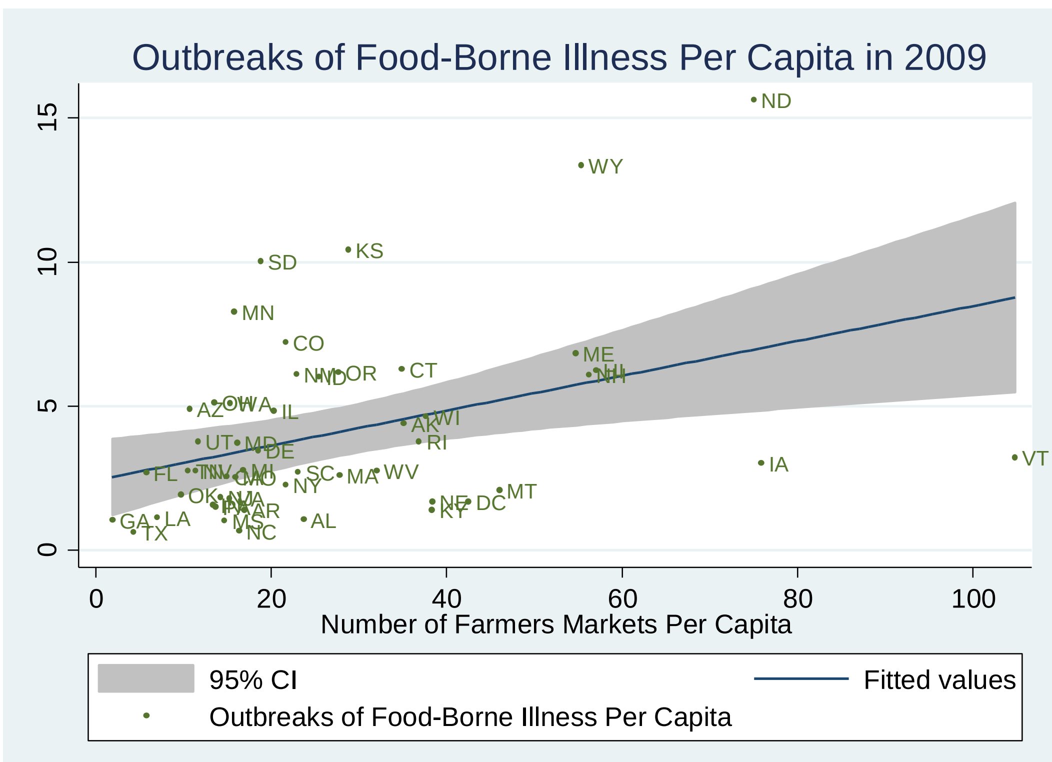


Figure 2d. Reported Outbreaks of Food-Borne Illness and Farmers Markets Per million in 2009.

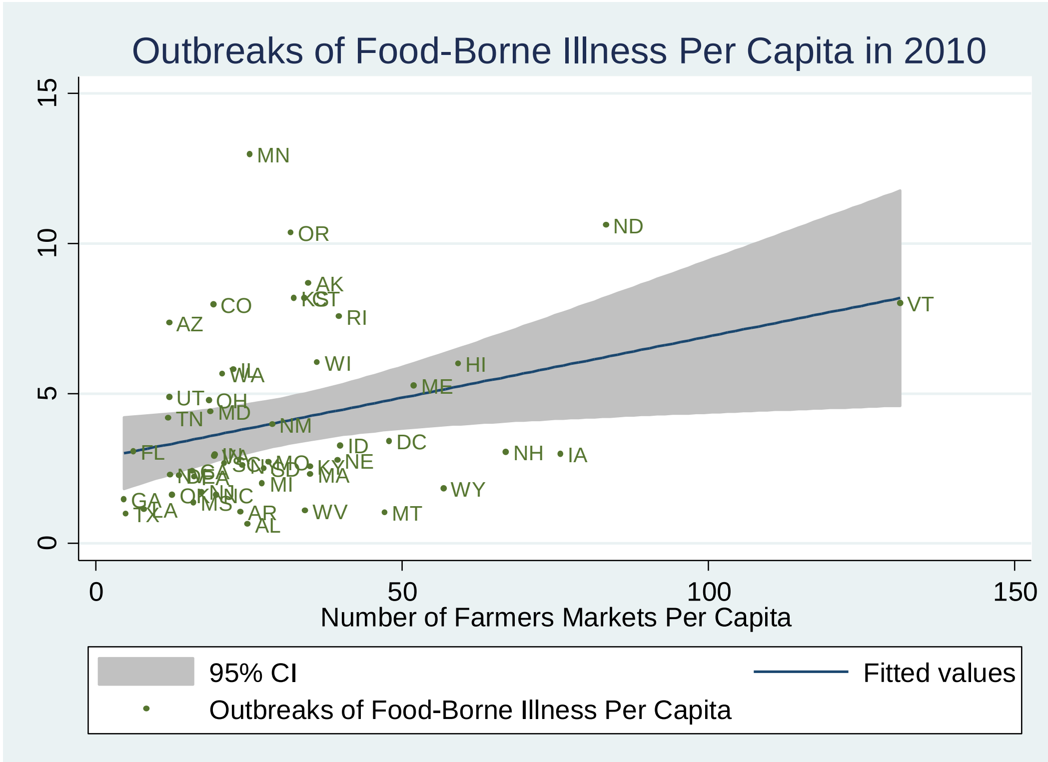


Figure 2e. Reported Outbreaks of Food-Borne Illness and Farmers Markets Per million in 2010.

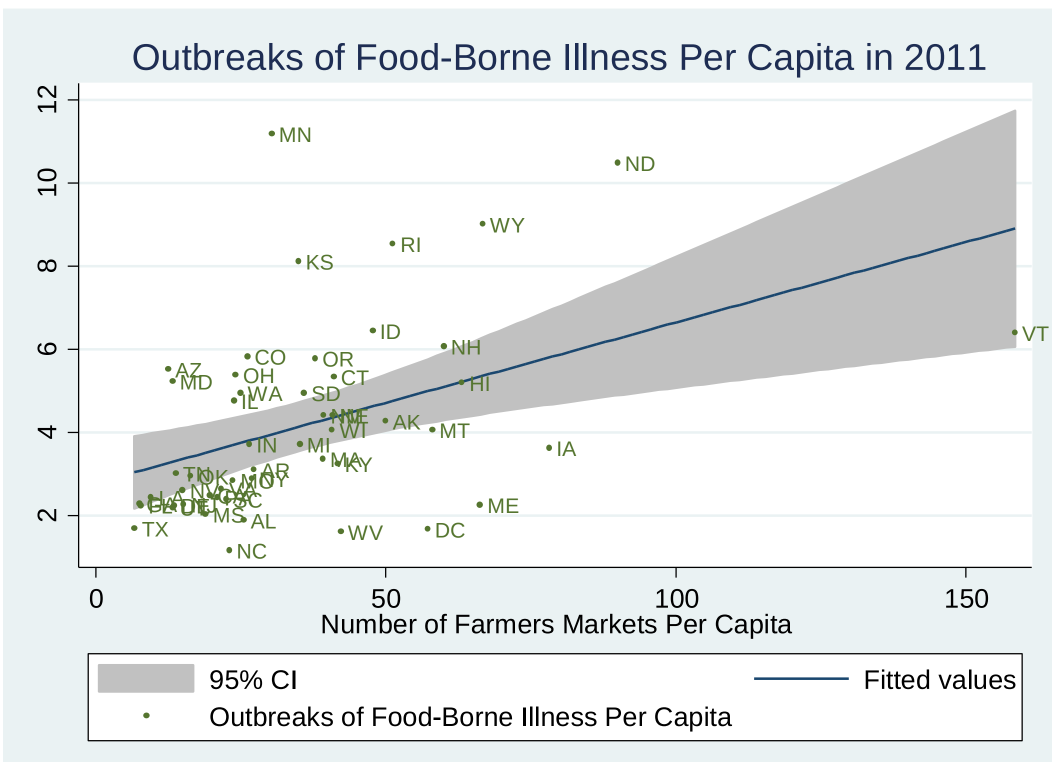


Figure 2f. Reported Outbreaks of Food-Borne Illness and Farmers Markets Per million in 2011.

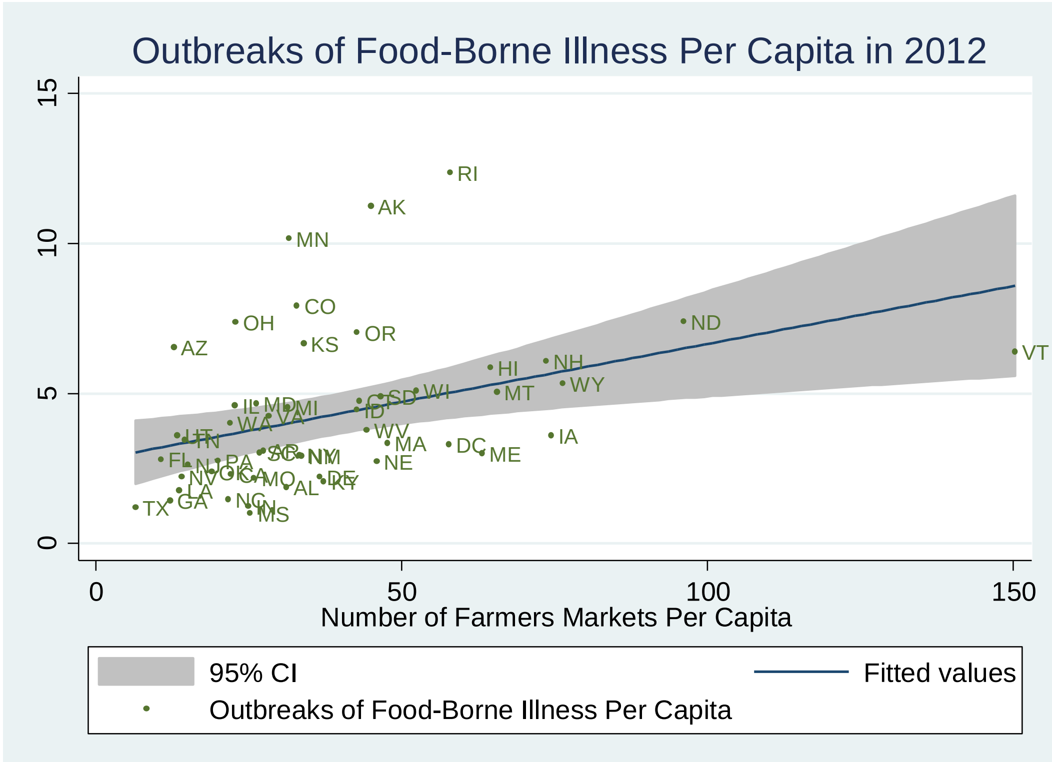


Figure 2g. Reported Outbreaks of Food-Borne Illness and Farmers Markets Per million in 2012.

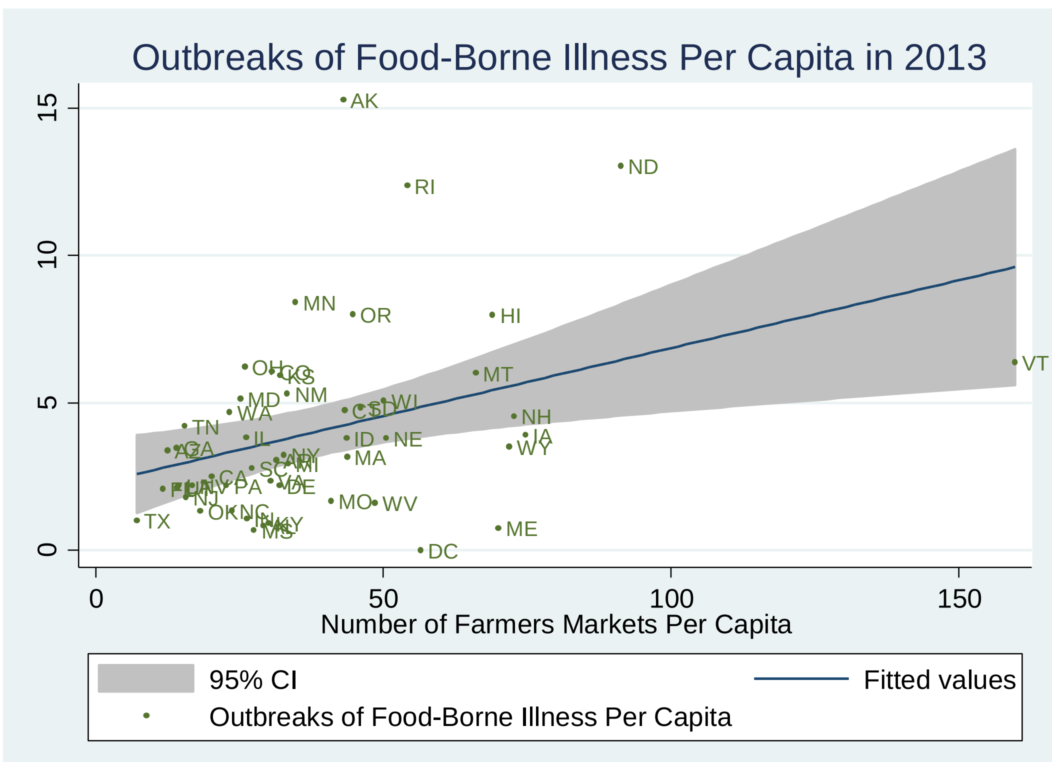


Figure 2h. Reported Outbreaks of Food-Borne Illness and Farmers Markets Per million in 2013.

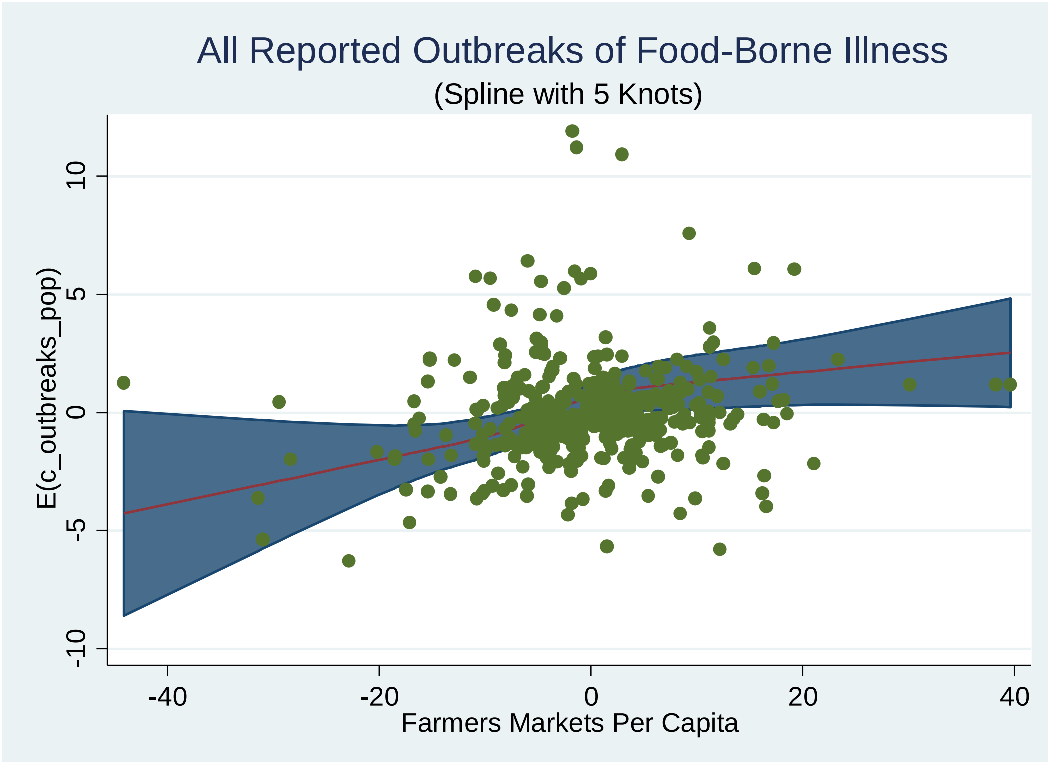
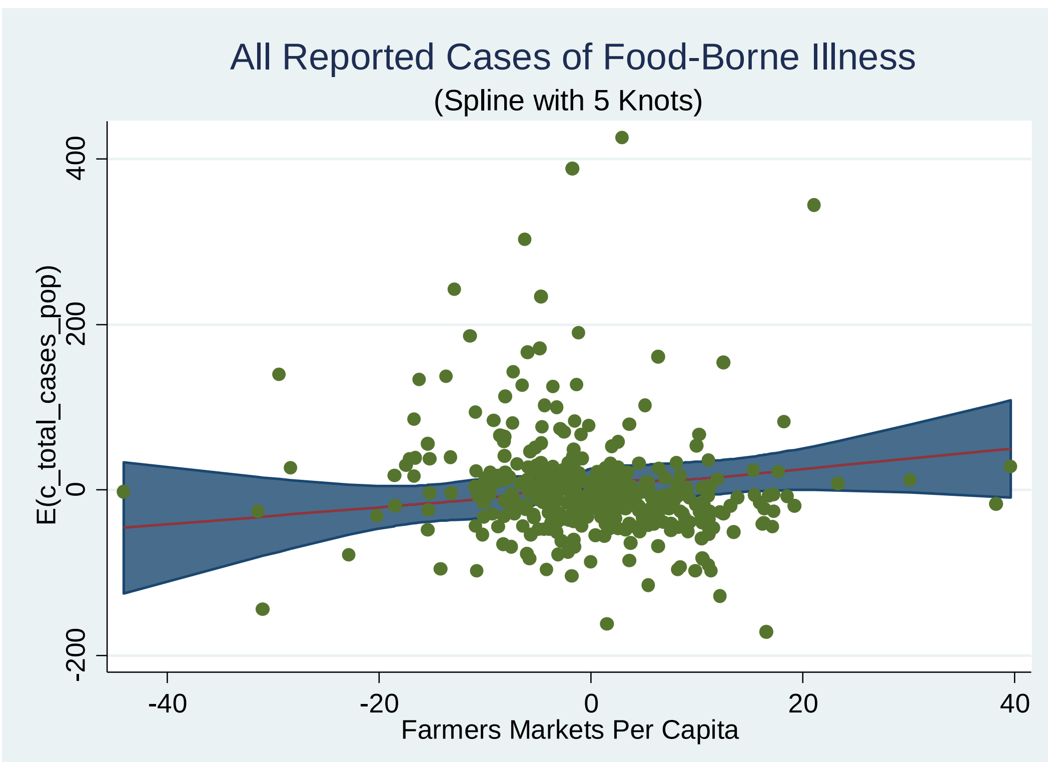


Figure 3a. Spline Regression for the Total Reported Number of Outbreaks of Food-Borne Illness



Figured 3b. Spline Regression for the Total Reported Number of Cases of Food-Borne Illness.

Table 1. Descriptive Statistics for All 50 States Plus the District of Colombia for the Period 2004, 2006, and 2008-2013 (n=408).

Variable	Mean	Std. Dev.
Reported Outbreaks		
Total	20.946	(25.838)
Norovirus	6.395	(10.517)
Salmonella Enterica	4.083	(3.900)
E. Coli Shiga	1.051	(1.500)
C. Perfringens	0.630	(1.149)
Campylobacter Jejuni	0.353	(0.737)
Scombroid	0.321	(0.960)
Staphylococcus Aureus	0.287	(0.924)
Reported Cases		
Total	354.417	(473.455)
Norovirus	157.880	(264.246)
Salmonella Enterica	51.944	(86.770)
C. Perfringens	27.115	(78.527)
E. Coli Shiga	6.657	(21.365)
Staphylococcus Aureus	5.034	(17.471)
Campylobacter Jejuni	9.370	(83.590)
Scombroid	1.066	(3.455)
Number of Farmers Markets	116.056	(112.787)
GDP (Millions of Dollars)	289,131.300	(353,223.100)
College Graduation Rate (Percent)	27.563	(5.582)
Number of Restaurants	11,270.670	(12,509.920)
Number of Bankruptcies	24,569.880	(30,672.320)
Population (Millions of Individuals)	5.951	(6.652)
Average Daily Minimum Temperature (Celsius)	5.460	(4.217)

Table 2. Ordinary Least Squares Estimation Results for the Total Number of Reported Outbreaks of Food-Borne Illness Per million.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	**Total**	**Norovirus**	**Salmonella**	**E. Coli**	**C. Perfringens**	**Campylobacter**	**Scombroid**	**Staph**
	Dependent Variable: Reported Outbreaks Per million.							
Farmers Markets Per million	0.089**	0.026**	0.008	-0.002	0.005	0.009***	0.013	0.004
	(0.034)	(0.012)	(0.006)	(0.003)	(0.004)	(0.003)	(0.011)	(0.003)
Multistate Outbreaks Not Recorded	1.102	0.279	-0.806**	-0.617***	0.118	0.032	0.317	0.175*
	(0.867)	(0.348)	(0.313)	(0.120)	(0.073)	(0.054)	(0.249)	(0.104)
GDP Per million	0.031	-0.088**	0.021	0.004	0.002	-0.004	-0.004	0.003*
	(0.055)	(0.034)	(0.024)	(0.004)	(0.006)	(0.010)	(0.004)	(0.002)
Proportion College Graduates	-0.371	-0.027	-0.076	-0.020	-0.038*	-0.011	-0.081	-0.028
	(0.337)	(0.136)	(0.077)	(0.018)	(0.022)	(0.020)	(0.075)	(0.040)
Restaurants Per million	-0.004	-0.003	-0.001	-0.001*	0.000	0.000	0.001	0.000
	(0.005)	(0.003)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Constant	18.503*	11.788*	3.498	2.206***	0.881	-0.311	-0.331	-0.014
	(10.053)	(6.629)	(2.177)	(0.749)	(0.703)	(0.789)	(0.440)	(0.379)
Observations	408	408	408	408	408	408	408	408
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.095	0.206	0.282	0.187	0.027	0.061	0.140	0.114

Standard errors clustered at the state level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3. Ordinary Least Squares Estimation Results for the Total Number of Reported Cases of Food-Borne Illness Per million

Variables	(1) Total	(2) Norovirus	(3) Salmonella	(4) E. Coli	(5) C. Perfringens	(6) Campylobacter	(7) Scombroid	(8) Staph
			Dependent Variable: Reported Cases Per million.					
Farmers Markets Per million	1.164**	0.774**	-0.076	-0.013	0.213	0.051*	0.032	0.094
	(0.515)	(0.337)	(0.131)	(0.019)	(0.194)	(0.029)	(0.024)	(0.081)
Multistate Outbreaks Not Recorded	24.351	21.335**	-12.762	-1.468**	5.085	1.438	0.775	2.733
	(15.367)	(9.502)	(9.238)	(0.639)	(6.947)	(2.096)	(0.516)	(1.863)
GDP Per million	-2.668**	-3.116**	-0.301	0.049	0.315	0.298	-0.011	0.042
	(1.320)	(1.352)	(0.438)	(0.068)	(0.361)	(0.376)	(0.011)	(0.044)
Proportion College Graduates	-3.705	1.733	-1.617	0.045	-1.571	-0.409	-0.181	-0.456
	(6.205)	(5.340)	(2.596)	(0.133)	(0.939)	(0.415)	(0.138)	(0.711)
Restaurants Per million	-0.107	-0.064	0.016	-0.003	0.003	-0.005	0.003	-0.001
	(0.105)	(0.090)	(0.022)	(0.002)	(0.021)	(0.008)	(0.002)	(0.002)
Constant	477.006**	226.242	50.514	3.705	25.194	5.848	-0.551	11.262
	(228.132)	(182.882)	(55.438)	(4.231)	(39.175)	(11.923)	(1.061)	(16.437)
Observations	408	408	408	408	408	408	408	408
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.145	0.186	0.056	0.033	0.020	0.034	0.102	0.036

Standard errors clustered at the state level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4. Robustness Checks for the Total Reported Number of Outbreaks of Food-Borne Illness

Variables	(1)	(2)	(3)	(4)	(5)
Dependent Variable: Reported Outbreaks of Food-Borne Illness Per million					
Farmers Markets Per million	0.098***	0.088***	0.069**	0.082*	0.147*
	(0.033)	(0.033)	(0.032)	(0.043)	(0.087)
Multistate Outbreaks Not Recorded	1.658	-0.178	-0.220	1.377	-0.575
	(1.025)	(0.398)	(0.513)	(1.520)	(0.369)
GDP Per million	0.033	0.018	0.053	0.053	0.059
	(0.054)	(0.052)	(0.107)	(0.065)	(0.056)
College Graduation Rate	-0.374	-0.261	-0.059	-0.156	-0.191
	(0.337)	(0.265)	(0.276)	(0.318)	(0.309)
Restaurants Per million	-0.004	-0.004	-0.009	-0.007	0.002
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
Farmers Markets Per million, Neighboring States	-0.001				
	(0.003)				
Constant	17.961*	393.739*	370.309	17.420	
	(10.109)	(217.411)	(498.770)	(11.227)	
Observations	408	408	408	408	384
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	No	No	No	No
Linear Trend	No	Yes	No	No	Yes
State-Specific Linear Trends	No	No	Yes	No	No
Regional Division-Year Fixed Effects	No	No	No	Yes	No
F-Statistic (Instrumental Variable)	-	-	-	-	17.16
R-squared	0.096	0.090	0.420	0.233	0.008
Number of state	51	51	51	51	48

Standard errors clustered at the state level in parentheses. The constant is omitted in column (5) because the Stata command used for the 2SLS results does not report a constant.

*** p<0.01, ** p<0.05, * p<0.1

Table 5. Robustness Checks for the Total Reported Number of Cases of Food-Borne Illness

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Dependent Variable: Reported Cases of Food-Borne Illness Per million					
Farmers Markets Per million	0.940	1.253**	2.366*	1.908*	5.834**
	(0.664)	(0.519)	(1.291)	(1.024)	(2.854)
Multistate Outbreaks Not Recorded	23.949	32.051**	35.139**	-21.609	21.825
	(15.388)	(14.472)	(13.695)	(24.314)	(14.229)
GDP Per million	-2.713**	-2.746*	-4.287**	-3.508***	-2.048
	(1.312)	(1.441)	(1.800)	(1.214)	(2.124)
College Graduation Rate	-3.613	-3.256	-0.058	0.305	-9.389
	(6.220)	(5.637)	(8.532)	(6.268)	(9.082)
Restaurants Per million	-0.108	-0.107	-0.048	-0.156*	0.208
	(0.107)	(0.103)	(0.148)	(0.092)	(0.186)
Farmers Markets Per million, Neighboring States	0.036				
	(0.052)				
Constant	477.032**	-1,355.173	1,419.708	527.776**	
	(236.552)	(5,765.531)	(6,840.285)	(219.608)	
Observations	408	408	408	408	384
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	No	No	No	No
Linear Trend	No	Yes	No	No	Yes
State-Specific Linear Trends	No	No	Yes	No	No
Regional Division-Year Fixed Effects	No	No	No	Yes	No
F-Statistic (Instrumental Variable)	-	-	-	-	17.16
R-squared	0.146	0.138	0.286	0.284	-0.094
Number of state	51	51	51	51	48

Standard errors clustered at the state level in parentheses. The constant is omitted in column (5) because the Stata command used for the 2SLS results does not report a constant.

*** p<0.01, ** p<0.05, * p<0.1

Table 6. Robustness Checks for the Number of Reported Outbreaks of Norovirus

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Dependent Variable: Reported Outbreaks of Norovirus Per million					
Farmers Markets Per million	0.034*	0.026**	0.033	0.020	0.141**
	(0.019)	(0.011)	(0.027)	(0.016)	(0.063)
Multistate Outbreaks Not Recorded	0.294	0.879***	0.816***	-0.810*	0.680***
	(0.332)	(0.187)	(0.216)	(0.429)	(0.164)
GDP Per million	-0.086**	-0.084**	-0.048	-0.079**	-0.060
	(0.032)	(0.034)	(0.067)	(0.032)	(0.044)
College Graduation Rate	-0.030	-0.060	0.033	0.075	-0.260
	(0.135)	(0.113)	(0.177)	(0.138)	(0.224)
Restaurants Per million	-0.003	-0.003	-0.005	-0.006**	0.006**
	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)
Farmers Markets Per million, Neighboring States	-0.001				
	(0.002)				
Constant	11.787*	-205.838	-75.691	15.464**	
	(6.284)	(128.813)	(217.037)	(6.300)	
Observations	408	408	408	408	384
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	No	No	No	No
Linear Trend	No	Yes	No	No	Yes
State-Specific Linear Trends	No	No	Yes	No	No
Regional Division-Year Fixed Effects	No	No	No	Yes	No
F-Statistic (Instrumental Variable)	-	-	-	-	17.16
R-squared	0.209	0.197	0.395	0.343	-0.312
Number of state	51	51	51	51	48

Standard errors clustered at the state level in parentheses. The constant is omitted in column (5) because the Stata command used for the 2SLS results does not report a constant.

*** p<0.01, ** p<0.05, * p<0.1

Table 7. Robustness Checks for the Reported Number of Reported Cases of Norovirus

Variables	(1)	(2)	(3)	(4)	(5)
Dependent Variable: Reported Cases of Norovirus Per million					
Farmers Markets Per million	0.634	0.894**	1.219	1.015*	6.735**
	(0.520)	(0.369)	(0.986)	(0.555)	(2.746)
Multistate Outbreaks Not Recorded	21.082**	30.307***	35.040***	-19.248	27.755***
	(9.458)	(6.733)	(8.533)	(13.483)	(9.014)
GDP Per million	-3.144**	-3.061**	-4.991**	-3.732***	-3.470
	(1.367)	(1.428)	(2.222)	(1.217)	(2.339)
College Graduation Rate	1.790	1.683	5.973	4.521	-8.795
	(5.321)	(4.463)	(5.769)	(5.467)	(8.097)
Restaurants Per million	-0.064	-0.068	-0.089	-0.107	0.272*
	(0.092)	(0.088)	(0.124)	(0.085)	(0.148)
Farmers Markets Per million,	0.023				
Neighboring States	(0.052)				
Constant	226.258	-4,015.134	-7,803.896*	294.961*	
	(188.627)	(4,269.200)	(4,477.680)	(162.838)	
Observations	408	408	408	408	384
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	No	No	No	No
Linear Trend	No	Yes	No	No	Yes
State-Specific Linear Trends	No	No	Yes	No	No
Regional Division-Year Fixed Effects	No	No	No	Yes	No
F-Statistic (Instrumental Variable)	-	-	-	-	17.16
R-squared	0.187	0.174	0.342	0.337	-0.492
Number of state	51	51	51	51	48

Standard errors clustered at the state level in parentheses. The constant is omitted in column (5) because the Stata command used for the 2SLS results does not report a constant.

*** p<0.01, ** p<0.05, * p<0.1

Table 8. Robustness Checks for the Number of Reported Outbreaks of Campylobacter

Variables	(1)	(2)	(3)	(4)	(5)
Dependent Variable: Reported Outbreaks of Campylobacter Per million					
Farmers Markets Per million	0.008*	0.008***	0.007	0.010***	0.010
	(0.005)	(0.002)	(0.005)	(0.003)	(0.018)
Multistate Outbreaks Not Recorded	0.031	0.055	0.053	0.115**	0.035
	(0.056)	(0.072)	(0.081)	(0.043)	(0.066)
GDP Per million	-0.005	-0.003	-0.003	-0.005	-0.010*
	(0.010)	(0.010)	(0.017)	(0.011)	(0.006)
College Graduation Rate	-0.011	-0.028**	-0.003	-0.012	-0.032
	(0.020)	(0.013)	(0.018)	(0.026)	(0.044)
Restaurants Per million	0.000	0.000	0.001	0.000	0.001
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Farmers Markets Per million,	0.000				
Neighboring States	(0.000)				
Constant	-0.311	-1.151	3.697	-0.599	
	(0.803)	(25.513)	(32.460)	(1.091)	
Observations	408	408	408	408	384
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	No	No	No	No
Linear Trend	No	Yes	No	No	Yes
State-Specific Linear Trends	No	No	Yes	No	No
Regional Division-Year Fixed Effects	No	No	No	Yes	No
F-Statistic (Instrumental Variable)	-	-	-	-	17.16
R-squared	0.062	0.040	0.096	0.175	0.027
Number of state	51	51	51	51	48

Standard errors clustered at the state level in parentheses. The constant is omitted in column (5) because the Stata command used for the 2SLS results does not report a constant.

*** p<0.01, ** p<0.05, * p<0.1

Table 9. Robustness Checks for the Number of Reported Cases of Campylobacter

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Dependent Variable: Reported Cases of Campylobacter Per million					
Farmers Markets Per million	0.031	0.035	0.062	0.023	0.042
	(0.042)	(0.039)	(0.067)	(0.046)	(0.144)
Multistate Outbreaks Not Recorded	1.402	1.962	0.074	1.550	0.198
	(2.088)	(1.995)	(1.391)	(1.097)	(0.664)
GDP Per million	0.294	0.329	1.202	0.436	-0.060
	(0.376)	(0.375)	(1.309)	(0.417)	(0.051)
College Graduation Rate	-0.401	-0.732	-0.409	-0.922	-0.391
	(0.414)	(0.453)	(0.721)	(0.642)	(0.443)
Restaurants Per million	-0.005	-0.003	-0.019	-0.016	0.001
	(0.009)	(0.008)	(0.023)	(0.015)	(0.005)
Farmers Markets Per million, Neighboring States	0.003				
	(0.004)				
Constant	5.850	167.672	2,814.133	35.311	
	(12.597)	(431.455)	(3,119.015)	(27.365)	
Observations	408	408	408	408	384
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	No	No	No	No
Linear Trend	No	Yes	No	No	Yes
State-Specific Linear Trends	No	No	Yes	No	No
Regional Division-Year Fixed Effects	No	No	No	Yes	No
F-Statistic (Instrumental Variable)	-	-	-	-	17.16
R-squared	0.034	0.024	0.101	0.198	0.012
Number of state	51	51	51	51	48

Standard errors clustered at the state level in parentheses. The constant is omitted in column (5) because the Stata command used for the 2SLS results does not report a constant.

*** p<0.01, ** p<0.05, * p<0.1

Table 10. Placebo and Falsification Tests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Placebo Tests						Falsification Test
	Total Reported		Norovirus		Campylobacter		
Variables	**Outbreaks**	**Cases**	**Outbreaks**	**Cases**	**Outbreaks**	**Cases**	**Bankruptcies**
Bankruptcies Per million	0.000	-0.000	0.000	-0.000	-0.000	0.000	-
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	
Farmers Markets Per million	-	-	-	-	-	-	-120.413
							(112.651)
Multistate Outbreaks Not Recorded	0.265	16.101	0.076	16.003	-0.011	0.908	4,686.795
	(0.767)	(16.684)	(0.478)	(12.970)	(0.047)	(1.908)	(3,279.139)
GDP Per million	0.070	-2.363**	-0.080**	-2.924**	-0.004	0.324	-380.827
	(0.060)	(1.094)	(0.030)	(1.181)	(0.011)	(0.393)	(230.338)
College Graduation Rate	-0.179	-2.315	0.013	2.593	-0.008	-0.279	-2,161.821
	(0.278)	(6.210)	(0.164)	(5.856)	(0.018)	(0.348)	(1,514.925)
Restaurants Per million	-0.006	-0.137	-0.004	-0.083	0.000	-0.006	-0.872
	(0.005)	(0.108)	(0.003)	(0.095)	(0.000)	(0.008)	(17.817)
Constant	18.400**	515.986**	12.308**	254.489	0.267	5.096	86,947.554**
	(8.905)	(200.087)	(5.334)	(154.743)	(0.830)	(11.152)	(36,116.654)
Observations	408	408	408	408	408	408	408
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.042	0.134	0.191	0.176	0.037	0.032	0.328

Standard errors clustered at the state level in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 11. Robustness Checks for Outliers and Leverage Points: Total Reported Number of Outbreaks of Food-Borne Illness

Variables	(1) Outliers	(2) Leverage Points	(3) Both
Dependent Variable: Reported Outbreaks of Food-Borne Illness Per million			
Farmers Markets Per million	0.076***	0.093**	0.079***
	(0.027)	(0.036)	(0.029)
Multistate Outbreaks Not Recorded	0.598	1.685	1.295
	(0.652)	(1.046)	(0.902)
GDP Per million	0.026	0.029	0.024
	(0.054)	(0.056)	(0.055)
College Graduation Rate	-0.124	-0.375	-0.128
	(0.247)	(0.337)	(0.245)
Restaurants Per million	-0.006	-0.004	-0.006
	(0.004)	(0.005)	(0.004)
Constant	15.512	17.748*	14.711
	(10.631)	(10.495)	(11.069)
Observations	407	407	406
State Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
R-squared	0.084	0.095	0.084

Standard errors clustered at the state level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 12. Robustness Checks for Outliers and Leverage Points: Total Reported Number of Cases of Food-Borne Illness

Variables	**(1) Outliers**	**(2) Leverage Points**	**(3) Both**
Dependent Variable: Reported Cases of Food-Borne Illness Per million			
Farmers Markets Per million	0.835*	1.119*	0.762
	(0.483)	(0.569)	(0.530)
Multistate Outbreaks Not Recorded	21.495	23.552	20.884
	(18.611)	(18.281)	(18.714)
GDP Per million	-2.449*	-2.645*	-2.412*
	(1.291)	(1.325)	(1.288)
College Graduation Rate	-4.161	-3.657	-4.086
	(6.343)	(6.206)	(6.339)
Restaurants Per million	-0.092	-0.109	-0.095
	(0.117)	(0.106)	(0.117)
Constant	458.624*	480.365**	463.005*
	(255.122)	(234.725)	(254.603)
Observations	407	407	406
State Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
R-squared	0.151	0.145	0.151

Standard errors clustered at the state level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix

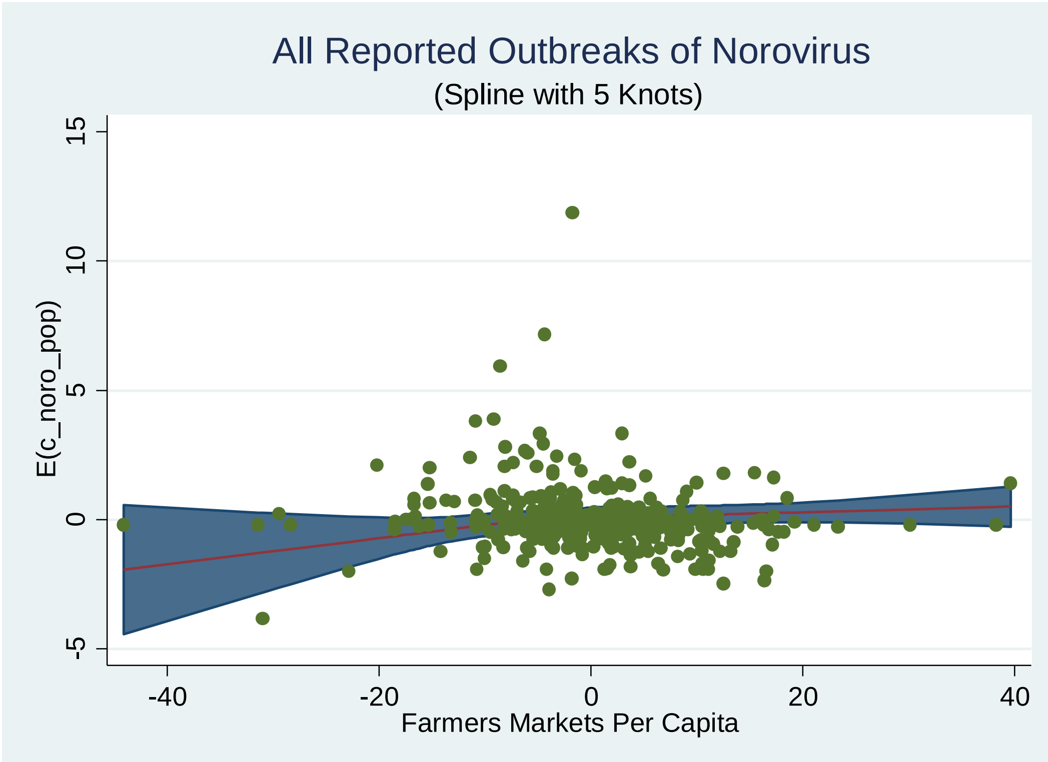


Figure A1. Spline Regression for the Total Reported Number of Outbreaks of Norovirus.

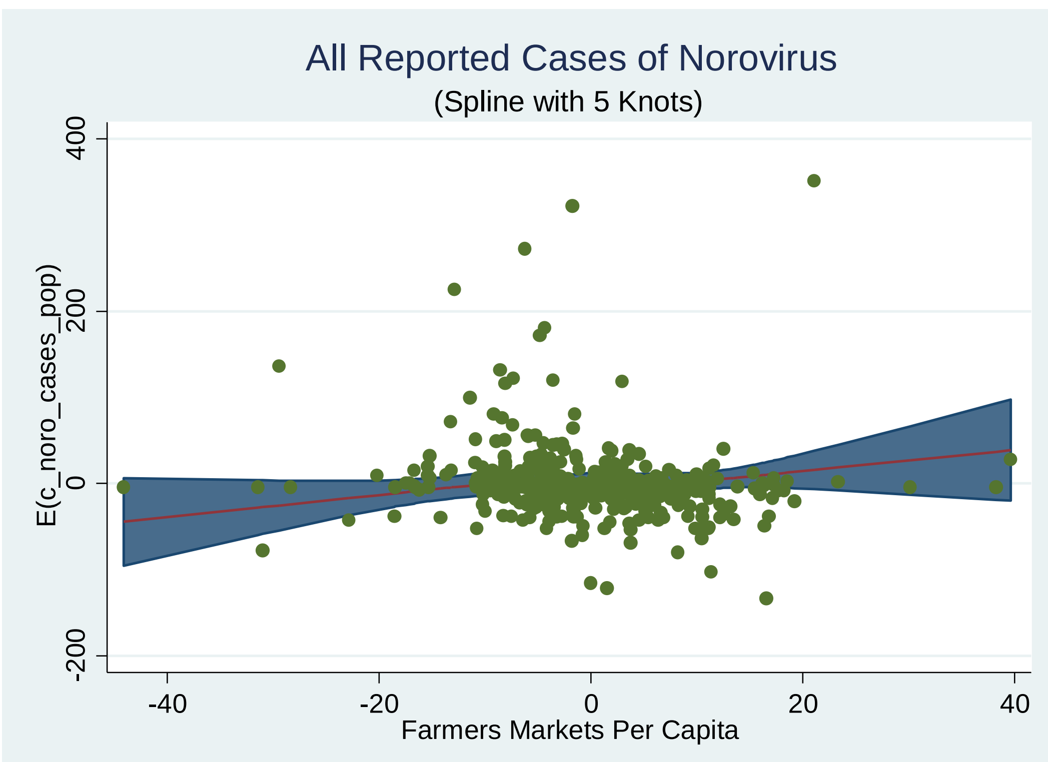


Figure A2. Spline Regression for the Total Reported Number of Cases of Norovirus.

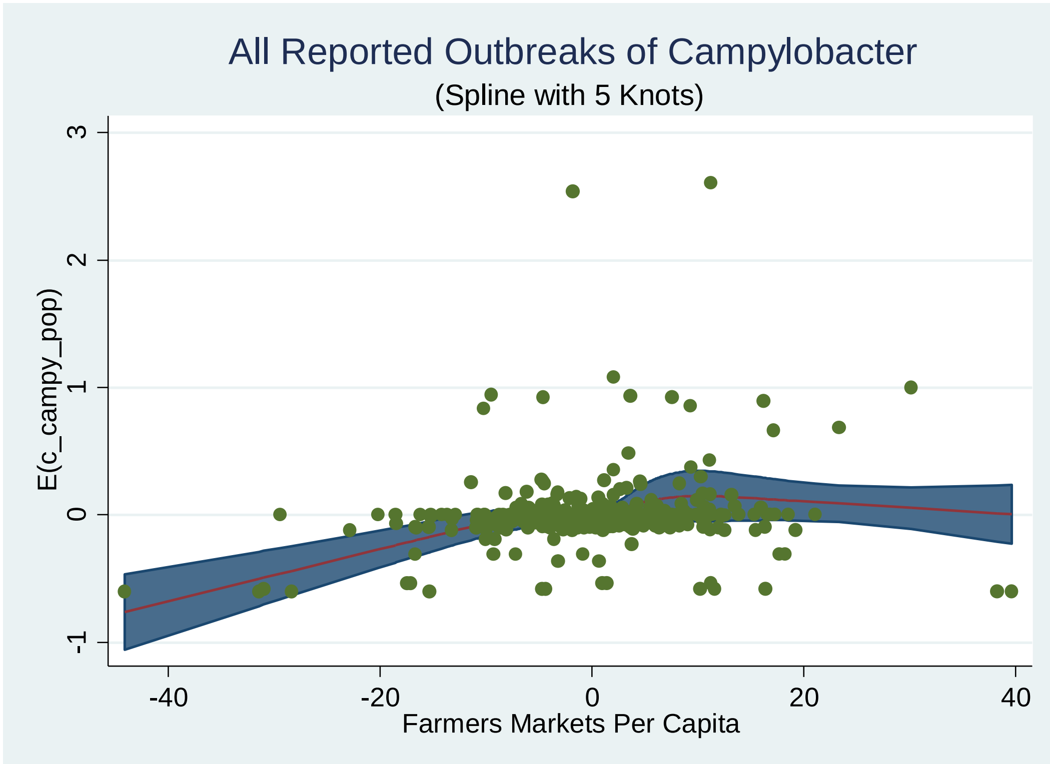


Figure A3. Spline Regression for the Total Reported Number of Outbreaks of Campylobacter.

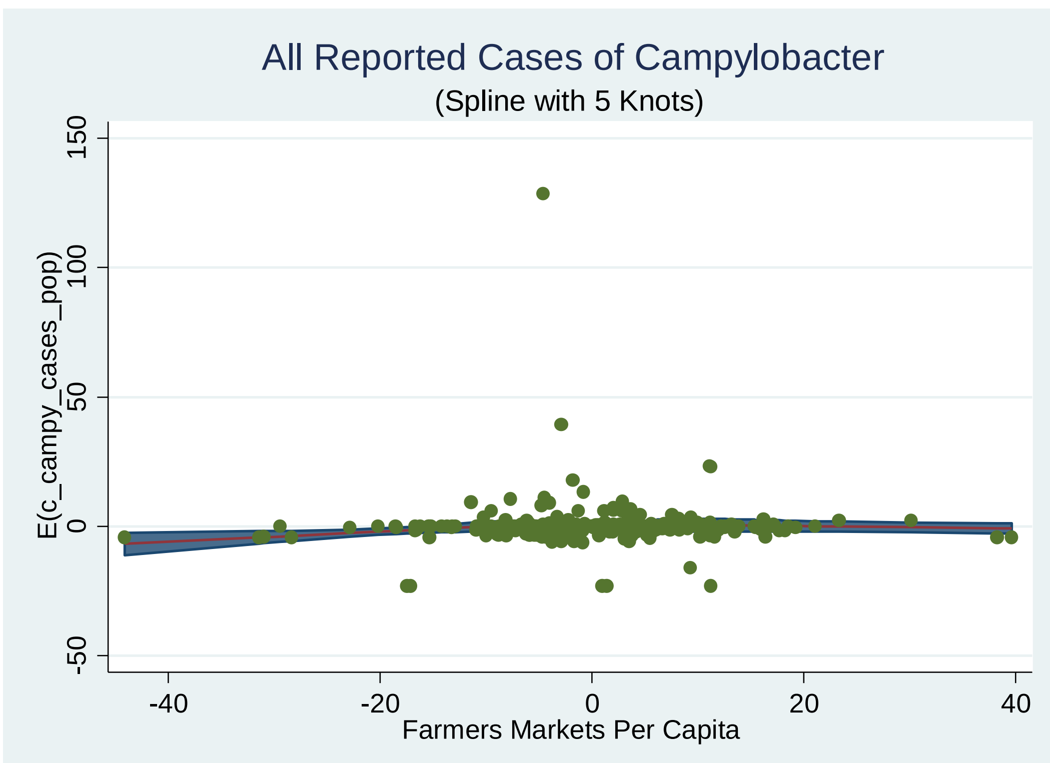


Figure A4. Spline Regression for the Total Reported Number of Cases of Campylobacter.

Table A1. Ordinary Least Squares Regression Results for the Total Number of Reported Outbreaks of Food-Borne Illness Per million

Variables	**(1)** **Total**	**(2)** **Norovirus**	**(3)** **Salmonella**	**(4)** **E. Coli**	**(5)** **C. Perfringens**	**(6)** **Campylobacter**	**(7)** **Scombroid**	**(8)** **Staph**
				Dependent Variable: Reported Outbreaks Per million.				
Farmers Markets Per million	0.089**	0.026**	0.008	-0.002	0.005	0.009***	0.013	0.004
	(0.034)	(0.012)	(0.006)	(0.003)	(0.004)	(0.003)	(0.011)	(0.003)
GDP Per million	0.031	-0.088**	0.021	0.004	0.002	-0.004	-0.004	0.003*
	(0.055)	(0.034)	(0.024)	(0.004)	(0.006)	(0.010)	(0.004)	(0.002)
Proportion College Graduates	-0.371	-0.027	-0.076	-0.020	-0.038*	-0.011	-0.081	-0.028
	(0.337)	(0.136)	(0.077)	(0.018)	(0.022)	(0.020)	(0.075)	(0.040)
Restaurants Per million	-0.004	-0.003	-0.001	-0.001*	0.000	0.000	0.001	0.000
	(0.005)	(0.003)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Constant	19.604*	12.067*	2.692	1.589**	1.000	-0.278	-0.014	0.161
	(9.897)	(6.475)	(2.115)	(0.689)	(0.680)	(0.807)	(0.492)	(0.452)
Observations	408	408	408	408	408	408	408	408
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.095	0.206	0.282	0.187	0.027	0.061	0.140	0.114

Standard errors clustered at the state level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A2. Ordinary Least Squares Regression Results for the Total Number of Reported Outbreaks of Food-Borne Illness Per million

Variables	(1) Total	(2) Norovirus	(3) Salmonella	(4) E. Coli	(5) C. Perfringens	(6) Campylobacter	(7) Scombroid	(8) Staph
Dependent Variable: Reported Outbreaks of Food-Borne Illness Per million.								
Farmers Markets Per million	0.104***	0.032**	0.005	-0.003	0.006	0.010***	0.016	0.005
	(0.026)	(0.014)	(0.007)	(0.004)	(0.004)	(0.003)	(0.010)	(0.003)
Farmers Markets Per million x	0.057**	0.026***	-0.012*	-0.002	0.001	0.003	0.011	0.003
Multistate Outbreaks Not Recorded	(0.023)	(0.009)	(0.007)	(0.002)	(0.003)	(0.003)	(0.008)	(0.003)
GDP Per million	0.077	-0.067***	0.012	0.002	0.003	-0.002	0.005	0.006*
	(0.047)	(0.025)	(0.031)	(0.004)	(0.007)	(0.009)	(0.006)	(0.003)
Proportion College Graduates	-0.376	-0.029	-0.075	-0.019	-0.038*	-0.011	-0.082	-0.028
	(0.306)	(0.131)	(0.086)	(0.018)	(0.022)	(0.020)	(0.071)	(0.039)
Restaurants Per million	-0.006	-0.004	-0.000	-0.001*	-0.000	0.000	0.001	0.000
	(0.004)	(0.003)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	20.833**	13.119**	2.434	1.024	1.024	-0.219	0.213	0.228
	(8.769)	(6.049)	(1.918)	(0.679)	(0.679)	(0.709)	(0.762)	(0.561)
Farmers Markets (Marginal Effect)	0.126***	0.042***	-0.002	-0.003	0.006	0.011***	0.020	0.006
	(0.032)	(0.015)	(0.015)	(0.004)	(0.005)	(0.003)	(0.013)	(0.004)
Observations	408	408	408	408	408	408	408	408
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.144	0.239	0.295	0.189	0.028	0.068	0.243	0.149

Standard errors clustered at the state level in parentheses

A dummy for whether multistate outbreaks were not recorded was among the regressors in each column but it was dropped in six out of eight regressions and is therefore not shown.

*** p<0.01, ** p<0.05, * p<0.1

Table A3. Ordinary Least Squares Regression Results for the Total Number of Reported Cases of Food-Borne Illness Per million

Variables	(1) Total	(2) Norovirus	(3) Salmonella	(4) E. Coli	(5) C. Perfringens	(6) Campylobacter	(7) Scombroid	(8) Staph
				Dependent Variable: Reported Cases Per million.				
Farmers Markets Per million	1.164**	0.774**	-0.076	-0.013	0.213	0.051*	0.032	0.094
	(0.515)	(0.337)	(0.131)	(0.019)	(0.194)	(0.029)	(0.024)	(0.081)
GDP Per million	-2.668**	-3.116**	-0.301	0.049	0.315	0.298	-0.011	0.042
	(1.320)	(1.352)	(0.438)	(0.068)	(0.361)	(0.376)	(0.011)	(0.044)
Proportion College Graduates	-3.705	1.733	-1.617	0.045	-1.571	-0.409	-0.181	-0.456
	(6.205)	(5.340)	(2.596)	(0.133)	(0.939)	(0.415)	(0.138)	(0.711)
Restaurants Per million	-0.107	-0.064	0.016	-0.003	0.003	-0.005	0.003	-0.001
	(0.105)	(0.090)	(0.022)	(0.002)	(0.021)	(0.008)	(0.002)	(0.002)
Constant	501.357**	247.576	37.752	2.236	30.279	7.286	0.223	13.994
	(223.626)	(179.739)	(54.649)	(4.081)	(37.236)	(12.724)	(1.064)	(17.886)
Observations	408	408	408	408	408	408	408	408
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.145	0.186	0.056	0.033	0.020	0.034	0.102	0.036

Standard errors clustered at the state level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A4. Ordinary Least Squares Regression Results for the Total Number of Reported Cases of Food-Borne Illness Per million

Variables	(1) Total	(2) Norovirus	(3) Salmonella	(4) E. Coli	(5) C. Perfringens	(6) Campylobacter	(7) Scombroid	(8) Staph
Dependent Variable: Reported Cases of Food-Borne Illness Per million.								
Farmers Markets Per million	1.486**	0.920**	-0.150	-0.017	0.267	0.072*	0.037	0.114
	(0.584)	(0.364)	(0.120)	(0.020)	(0.195)	(0.042)	(0.022)	(0.075)
Farmers Markets Per million x	1.216*	0.550*	-0.279	-0.017	0.204	0.079	0.019	0.074
Multistate Outbreaks Not Recorded	(0.685)	(0.282)	(0.203)	(0.030)	(0.186)	(0.075)	(0.015)	(0.054)
GDP Per million	-1.707	-2.681*	-0.522	0.035	0.476	0.361	0.004	0.100
	(1.468)	(1.490)	(0.621)	(0.052)	(0.404)	(0.410)	(0.012)	(0.081)
Proportion College Graduates	-3.820	1.680	-1.590	0.046	-1.591	-0.417	-0.182	-0.463
	(5.873)	(5.316)	(2.767)	(0.132)	(1.000)	(0.415)	(0.129)	(0.681)
Restaurants Per million	-0.156	-0.085	0.027	-0.002	-0.005	-0.008	0.002*	-0.004
	(0.104)	(0.094)	(0.026)	(0.002)	(0.021)	(0.010)	(0.001)	(0.005)
Constant	527.438**	257.292	31.771	1.862	34.653	8.991	0.640	14.826
	(201.596)	(176.233)	(48.455)	(4.315)	(39.016)	(11.710)	(1.484)	(22.693)
Farmers Markets (Marginal Effect)	1.942***	1.126***	-0.255	-0.024	0.343	0.102*	0.044	0.142
	(0.761)	(0.404)	(0.167)	(0.021)	(0.235)	(0.051)	(0.027)	(0.094)
Observations	408	408	408	408	408	408	408	408
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.171	0.197	0.068	0.034	0.027	0.041	0.143	0.049

Standard errors clustered at the state level in parentheses

A dummy for whether multistate outbreaks were not recorded was among the regressors in each column but it was dropped in all eight regressions and is therefore not shown.

*** p<0.01, ** p<0.05, * p<0.1

Table A5. Robustness Checks for Outliers and Leverage Points: Reported Number of Outbreaks of Norovirus

Variables	(1) Outliers	(2) Leverage Points	(3) Both
Dependent Variable: Reported Outbreaks of Norovirus Per million			
Farmers Markets Per million	0.019*	0.023*	0.016
	(0.011)	(0.014)	(0.012)
Multistate Data Not Recorded	0.902***	-0.277	0.879**
	(0.329)	(0.515)	(0.383)
GDP Per million	-0.019	-0.086**	-0.018
	(0.028)	(0.034)	(0.029)
College Graduation Rate	-0.018	-0.024	-0.015
	(0.150)	(0.136)	(0.149)
Restaurants Per million	0.002*	-0.003	0.002
	(0.001)	(0.003)	(0.001)
Constant	-3.013	12.494*	-2.838
	(3.527)	(6.848)	(3.734)
Observations	407	407	406
State Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
R-squared	0.140	0.205	0.138

Standard errors clustered at the state level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A6. Robustness Checks for Outliers and Leverage Points: Reported Number of Cases of Norovirus

Variables	(1) Outliers	(2) Leverage Points	(3) Both
Dependent Variable: Reported Outbreaks Per million			
Farmers Markets Per million	0.605*	0.726*	0.558
	(0.333)	(0.373)	(0.358)
Multistate Data Not Recorded	37.593***	6.686	36.901***
	(8.427)	(14.657)	(10.983)
GDP Per million	-1.316	-3.091**	-1.292
	(1.254)	(1.356)	(1.264)
College Graduation Rate	1.966	1.783	2.015
	(5.709)	(5.347)	(5.711)
Restaurants Per million	0.085	-0.066	0.083
	(0.057)	(0.091)	(0.057)
Constant	-160.464	243.594	-157.029
	(102.475)	(188.129)	(106.415)
Observations	407	407	406
State Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
R-squared	0.147	0.197	0.147

Standard errors clustered at the state level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A7. Robustness Checks for Outliers and Leverage Points: Reported Number of Outbreaks of Campylobacter

Variables	(1) Outliers	(2) Leverage Points	(3) Both
Dependent Variable: Reported Outbreaks Per million			
Farmers Markets Per million	0.007***	0.011***	0.008***
	(0.002)	(0.004)	(0.003)
Multistate Data Not Recorded	0.023	0.053	0.034
	(0.054)	(0.059)	(0.056)
GDP Per million	-0.005	-0.006	-0.005
	(0.010)	(0.010)	(0.010)
College Graduation Rate	-0.004	-0.014	-0.005
	(0.019)	(0.020)	(0.019)
Restaurants Per million	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Constant	-0.646	-0.461	-0.713
	(0.736)	(0.728)	(0.741)
Observations	407	407	406
State Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
R-squared	0.060	0.079	0.069

Standard errors clustered at the state level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A8. Robustness Checks for Outliers and Leverage Points: Reported Number of Cases of Campylobacter

Variables	(1) Outliers	(2) Leverage Points	(3) Both
Dependent Variable: Reported Outbreaks Per million			
Farmers Markets Per million	0.033	0.065*	0.035
	(0.026)	(0.038)	(0.030)
Multistate Data Not Recorded	1.354	1.547	1.373
	(2.105)	(2.078)	(2.079)
GDP Per million	0.297	0.291	0.296
	(0.376)	(0.377)	(0.378)
College Graduation Rate	-0.342	-0.424	-0.345
	(0.411)	(0.412)	(0.408)
Restaurants Per million	-0.004	-0.004	-0.004
	(0.008)	(0.009)	(0.009)
Constant	2.795	5.035	2.669
	(12.306)	(11.892)	(12.386)
Observations	407	407	406
State Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
R-squared	0.034	0.034	0.034

Standard errors clustered at the state level in parentheses
*** p<0.01, ** p<0.05, * p<0.1