Identifying and Measuring the Effect of Firm Clusters Among Certified Organic Processors and Handlers

Edward C. Jaenicke
Associate Professor
Department of Agricultural Economics and Rural Sociology
Penn State University
208A Armsby
University Park, PA 16802
Tel: (814) 865-5282
Email: tjaenicke@psu.edu

Stephan J. Goetz
Professor of Agricultural and Regional Economics, Penn State University and
Director of the Northeast Regional Center for Rural Development

Ping-Chao Wu
Graduate Research Assistant, Penn State University

Carolyn Dimitri
USDA/Economic Research Service

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association, Milwaukee, Wisconsin, July 26–28, 2009.

This project was supported by the Northeast Regional Center for Rural Development and a Cooperative Agreement with the USDA's Economic Research Service

Introduction

Despite the explosion of interest in industry clusters, formal studies evaluating the impacts of firm agglomeration on industry performance remain relatively scarce. Important exceptions include Gibbs and Bernat (1997) and, more recently, Gabe (2004 and 2008) and Graham and Kim (2008). Previous research has focused primarily on identifying and measuring industry clusters (e.g., Porter, 2003, Goetz, Shields, and Wang, 2008), or on assessing factors underlying their formation (Ellison and Glaeser, 1997; Ellison, Glaeser, and Kerr, 2007). In the food and agricultural sector, Roe, Irwin, and Sharp (2002) and Davis and Schluter (2005) directly or indirectly quantify agglomerations effects in spatial hog production trends and new food manufacturing investments. However, considering that Porter (1998) used the California wine industry in his seminal work, the rarity of empirical work is especially glaring in the food and agricultural sector, and even more so for organic produce, which despite a growing importance is itself an under-researched component of the food system. The agricultural sector is particularly relevant for agglomeration research because food production and distribution are closely tied to space. In other words, the consequences of agglomerations in which buying and selling firms both compete and cooperate with one another as a result of proximate locations may be especially critical for food and agricultural firms.

In addition to the issue of measuring agglomeration impacts, important questions remain about the definition and measurement of clusters in agriculture and in other industries, and how these affect performance measures. Unresolved and largely unstudied is just how many firms and what geographic space should define a cluster, or how this cluster variable should be specified. Gabe (2004), for example, follows a line of research that measures industry agglomeration as an industry intensity variable and compares a region's particular industry concentration with the national average. In addition, Gabe (2004) investigates the implications of measuring this intensity variable on a county or municipality level. Alternatively, Cainelli (2008) measures agglomeration as a binary indicator variable that takes the value of 1 if a firm belongs to an officially defined Italian business district. This almost purely empirical issue of cluster identification could be expected to a strong factor when measuring the impact of clusters in the organic industry. In our study, the agglomeration measure is a binary variable (like Cainelli 2008) that takes the value 1 if some minimum number of similar firms is present within a specific geographic area. Unlike other studies, however, we investigate the implications of agglomeration impacts from specifying alternative minimum numbers of firms within an industry cluster.

Economic Clusters in the Organics Industry

This study targets the certified organic "handling" sector, which lies between production and retailing. Organic handlers are firms that serve as packers, shippers, manufacturers, processors, or brokers, distributors, and wholesalers. According to Dimitri and Oberholtzer (2008), markets in this industry grew rapidly, increasing 17 percent a year between 1995 and 2006. While this growth reflects the potential for increased profits in the organic handling sector, it may simultaneously lead to industry growing pains as supply chains continually shift to accommodate more organic production and consumer demand. Within this shifting supply chain (or market channel) in which organic firms face the competitive challenges of a high-growth sector, firm clusters could have a positive or negative impact on firm performance depending on whether cooperative or competitive forces dominate.

Very little empirical research is available to guide our operating definition of industry clusters among organic handlers. Figure 1, reproduced from Dimitri and Oberholtzer (2008, p. 12) shows the geographic dispersion of certified organic handlers. The map uses U.S. zip codes to cluster firms, and the number of organic firms within a single zip code ranges from zero to over ten. Our preliminary results showed little difference from using zip codes or counties, but more differences when the agglomeration variable varies with the number of firms in a cluster. More specifically, we use U.S. counties as the cluster boundaries and explicitly investigate how the number of firms within a boundary affects the estimated impacts of a firm cluster.

Our effort to explore various definitions of firm clusters is secondary, however, to our primary goal of quantifying the impact of a cluster on firm decisions or firm performance. Using data from a population survey of U.S. certified organic handlers (Dimitri and Oberholtzer 2008), we investigate the impact of clustering on several firm-level variables that reflect firms' performance and their marketing or procurement decisions. These firm-level variables include the following: total gross sales per employee, total gross sales, total number of full-time employees, the percentage of handlers' sales total sales that is organic, the percentage of handlers' total procurement that is organic, and the percentage of organic products sold or procured locally, regionally, nationally, or internationally.

Before we can estimate the impact of clusters on these firm-level variables, our empirical investigation has several preliminary steps. The first step is to operationalize the definition of a firm cluster. Here, we find little difference from using zip codes or counties; we do, however, find important differences when the definition of a cluster depends on the number of firms in a cluster.[1] Our second step is to account for potential endogeneity in the cluster variable by estimating a cluster-formation equation where the formation or presence of a cluster, from the firm's perspective, is a binary dependent variable. Our third step is to estimate the impact of a cluster's presence, along with other exogenous factors, on the firm-level output variables mentioned above. Because cluster impacts are conditional on cluster formation, which is itself endogenous, we model cluster formation as a treatment effect and estimate the second and third steps simultaneously following the maximum-likelihood methods outlined in Cameron and Trivedi (2005, Chapter 25). Our last step is to replicate these system estimations after varying the minimum number of firms that define a cluster.

Our results confirm that the presence of a firm cluster often does have a significant impact on firm-level performance or decision variables. For example, clustered firms have more than $1 million in additional sales per-employee. In addition, the results from our last step show that the impact of clusters on firm performance and other firm decisions is sensitive to the minimum number of firms chosen to define a cluster. When a firm cluster is defined as a three or more organic firms located within a county, for example, clusters positively impact a firm's total gross sales. However, when the cluster is defined using a larger minimum number of firms (e.g., nine or more organic firms within a county), then a cluster's presence negatively impacts a firm's total sales.

[1] We eventually present estimated econometric models using a county-based definition of a cluster because more socio-economic data are available at the county level than the zip code level.

These results and others, along with some robustness checks, are presented and discussed following a more formal presentation of our methods and data.

Model and Methods

In this section, we develop an econometric model where an equation that describes cluster formation is linked to an equation that describes the impact of clusters on firm performance or firm decisions. Our model characterizes a cluster's impacts as a treatment effect (see for example Cameron and Trivedi, 2005), where cluster formation is endogenous and therefore its effect on firm performance may be subject to selection bias. Modeling cluster impacts first, we let y^j_i denote firm-level decision j of firm i or an indicator of firm i's performance, and we allow this variable to depend on the presence of a firm cluster, C_n, and other controlling factors, $\boldsymbol{x}$, so that

(1) $$y^j_i = \alpha C_{n,i} + \boldsymbol{x}_i'\boldsymbol{\beta}_1 + \varepsilon_{1i} ,$$

where $\boldsymbol{\beta}_1$ are the estimated coefficients on the controlling factors, $\boldsymbol{x}'\boldsymbol{\beta}_1$ takes a linear form by assumption, ε_{1i} is an error term described below, and α is the impact of firm clustering on y_j. In the estimated models that follow, we assume that j takes on a number of forms to reflect J different firm-level decisions or performance measures.[2] Controls in the $\boldsymbol{x}$ vector are variables that describe the function of firm i or the demographic surroundings of firm i. Described this way, (1) is not intended to represent a structural equation that describes a firm's optimizing behavior. Rather, it explains variations in observed differences in performance measures or output variables that might come from optimizing behavior.[3]

[2] In most of the following discussion, the index j will be suppressed to reduce the notational burden.
[3] While equation (1) is admittedly ad hoc in nature, it could be thought of as analogous to an output supply equation obtained by applying Hotelling's lemma to a restricted profit function of the form $\pi(\boldsymbol{w}, p; C_n, \boldsymbol{x}, \boldsymbol{z})$, where $\boldsymbol{w}$ and p are input and output prices, and C_n, $\boldsymbol{x}$, and $\boldsymbol{z}$ are treated as fixed factors. To make the

The binary cluster variable relevant to firm i, $C_{n,i}$, is modeled as the outcome of an unobserved latent variable, $C^*_{n,i}$. Both the observed and latent variables are indexed by n to imply that the definition of a cluster depends on the minimum number of firms defined to make up a cluster. We assume that $C^*_{n,i}$ is a linear function of a second set of controlling factors, $\boldsymbol{z}$, which are based on Goetz (1997) and some of which may overlap with $\boldsymbol{x}$, so that

(2) $$C^*_{n,i} = \boldsymbol{z}_i'\boldsymbol{\beta}_2 + \varepsilon_{2i} ,$$

and the observed cluster variable is

(3) $$C_{n,i} = \begin{cases} 1, & if\ \ C^*_{n,i} > 0 \\ 0, & otherwise \end{cases} .$$

Because of the endogeneity of $C_{n,i}$ and the possibility that cluster formation and cluster impacts may occur simultaneously, equations (1) and (3) are estimated jointly. The error terms ε_{2i} and ε_{2i} are bivariate normal with mean zero and covariance matrix

$$\begin{bmatrix} \sigma & \rho \\ \rho & 1 \end{bmatrix} .$$

Given this specification, the log likelihood for observation i is:

(4) $$\ln L = \begin{cases} ln\Phi\left\{\frac{z_i'\beta_2+(y_i-x_i'\beta_1-\alpha)\rho/\sigma}{\sqrt{1-\rho^2}}\right\} - \frac{1}{2}\left(\frac{y_i-x_i'\beta_1-\alpha}{\sigma}\right)^2 - \ln\left(\sqrt{2\pi}\sigma\right), & y_{\boldsymbol{i}} = 1 \\ ln\Phi\left\{\frac{-z_i'\beta_2-(y_i-x_i'\beta_1)\rho/\sigma}{\sqrt{1-\rho^2}}\right\} - \frac{1}{2}\left(\frac{y_i-x_i'\beta_1}{\sigma}\right)^2 - \ln\left(\sqrt{2\pi}\sigma\right), & y_{\boldsymbol{i}} = 0 \end{cases} ,$$

where $\Phi(.)$ is the cumulative standard normal distribution function.

To examine the effect of firm clustering on the dependent variable y, we are most interested in estimating parameter α. However, as Cameron and Trivedi (2005) explain for more general cases, the average effect of a firm cluster must take into account the

---

analogy to an output supply equation complete, $\boldsymbol{w}$ and p must vary proportionally across firms at any point in time, and must be estimated with cross-sectional data. In that case, the proportional prices would be incorporated into an estimated constant term.

endogeneity of clusters. In other words, the average effect must account for potential selection bias and therefore, parameter ρ as well. In our case, the average effect is the difference in the dependent variable conditional on whether a firm is in a cluster or not. Greene (2008, p. 890) shows that this average effect of a cluster on the dependent variable is:

$$(5) \qquad E(y_i|C_{ni}=1) - E(y_i|C_{ni}=0) = \alpha + \rho\sigma\left[\frac{\phi(\boldsymbol{z_i}'\boldsymbol{\beta}_2)}{\Phi(\boldsymbol{z_i}'\boldsymbol{\beta}_2)\{1-\Phi(\boldsymbol{z_i}'\boldsymbol{\beta}_2)\}}\right],$$

where $\phi(.)$ is the standard normal density function.[4] From (5), one can see that α would provide an appropriate estimate of the cluster effect if $\rho = 0$, which occurs if the cluster formation and the cluster effect equations are independent. On the other hand, if ρ is positive, then α would underestimate the cluster effect.

Established theories on agglomeration and regional development provide little guidance for identifying elements of $\boldsymbol{x}$, $\boldsymbol{z}$, or even the best choices of for the dependent variable, y^j_i. In prior empirical research, two choices for the dependent variable include the change in output (Cainelli 2008), and investment per worker (Gabe 2004). In our study, y^j_i represents total gross sales, total employees, total sales per employee, and ten other firm-level variables available from the dataset described in the next section.

Choices for elements of $\boldsymbol{x}$ and $\boldsymbol{z}$ can be more problematic. For research based on the estimation of structural equations (e.g., Graham and Kim 2008, and Cainelli 2008), the choice are somewhat clear. For non-structural model estimation, however, there are more choices. Roe, Irwin, and Sharp (2002) list seven categories of variables in their estimatable model: (i) the agglomeration variable, (ii) urban encroachment and

[4] The statistical software package Stata (release 10.0), which is later used for estimation, describes the same formula but also allows one to estimate the left-hand side of equation (5) directly by recovering the predicted values of y, conditional on the cluster variable being equal to zero or one.

population characteristic variables, (iii) input availability variables, (iv) firm productivity and specialization variables, (v) local economic variables, (vi) market access variables, and (vii) regulatory variables. Our study uses the first five categories, (i), (ii), (iii), (iv), and (v), to help identify available data that can be used for elements of $\boldsymbol{x}$ and $\boldsymbol{y}$. These categories are identified on the list of variables found in Table 1.

Data

Much of our data comes from a 2004 survey of certified organic handlers administered by USDA's Economic Research Service. Dimitri and Oberholtzer (2008) describe in detail the survey methodology and results. The survey included questions on firm characteristics as well as marketing and procurement practices. For each firm surveyed, we used county codes to identify firm clusters. For example, using one definition, firm i was said to be in a cluster if at least two additional certified organic handlers from the survey were located in the same county. In this example, $n = 3$, so if firm i is part of this cluster, then $C_{3,i} = 1$.

In addition to survey data, we also collected data from the U.S. Census of Agriculture to help describe the economic conditions found in an individual firm's county. Table 1 lists and describes the variables from both the USDA survey and Census data used in our analysis. In total, 316 firms in the survey have a complete set of data for all the variables listed in Table 1.[5] Several of the variables listed in Table 1 are the result of minor manipulations of the original handler survey data. For example, in the original survey, organic handlers were asked to prioritize the "Availability of year-round supply"

[5] The results that follow fix the sample size at 316, the minimum sample size where valid observations for variables used in all the estimated models. Because survey response varied across the 13 questions that generate the different dependent variables used in these models, an alternative approach is to let the sample size "float" for each of the estimated models. When we used this approach, we found that the results were not substantially different from the results presented here.

and were given four choices: High Priority, Medium Priority, Low Priority, and No Difference. Responses to this question were converted to a binary variable that equals 1 if the firm answered High Priority or Medium Priority, and 0 otherwise. In the case of one question in particular, Total Gross Sales, we imposed a substantial change to the raw data. The sales question in the USDA survey asked organic handlers to describe their total annual gross sales by picking from among a list of seven sales-range categories. In our current analysis, we transform this categorical variable to an integer by using the midpoints of the sales categories. The conversion for the highest sales category (over $100 million) is chosen as 1.5 times the cutoff ($150 million). We also created a productivity or output efficiency variable by dividing this new total sales variable by a firm's number of full-time employees. Much of the survey-based data in Table 1 is described in Dimitri and Oberholtzer (2008), although our smaller sample size may lead to some discrepancies in mean values. Additional county-level data comes from the Census of Agriculture and the Bureau of Economic Analysis.

We identify 31 variables from the organic handler survey that might be expected to influence firms' decisions or performance and help control for impacts not due to the presence of a firm cluster. These variables form the basis for the $\boldsymbol{x}$ and $\boldsymbol{z}$ vectors.[6] In all cases, these variables fit within categories used by Roe, Irwin, and Sharp (2002). Table 1 identifies each variable with a particular category.

Upon examination of Table 1, one can see that more than half of the 316 firms are part of clusters if we use the C_3 definition. Alternatively, only 21.5 percent of the total lie in a cluster if 10 is chosen as the minimum number of clustering firms (C_{10i}). Organic

[6] Greene (2008) notes that joint ML estimation of the system can be complicated by identification issues. For this reason, we are careful to include some elements of $\boldsymbol{x}$ and $\boldsymbol{z}$ that are unique to each vector.

sales are mostly national (38.7 percent) or regional (29.8 percent); organic procurement, however, is split almost evenly among local, regional, national and international sources. Table 1 also shows that most firms function as manufacturers or processors, and the average is around 71.8 percent. While 16.8 percent of the total sample handle organic manufactured products related to grain or oilseed milling, only 5.1 percent of the total relate to animal slaughtering or processing. More than 60 percent of the firms in the data place a high or medium priority on procuring supplies locally, and more than 70 percent of the total use contracts for procurement.

Results and Discussion

Thirteen dependent variables listed in Table 1, along with each of the eight cluster variables, are estimated in Maximum Likelihood systems represented by equations (1), (2), and (3). Because reporting 104 separate ML estimation results is impractical, we present instead a small illustrative selection of results.[7] First, we select one particular cluster variable, C_6, chosen because its criterion of requiring at least six organic firms to define a cluster is in the middle of our range of examined definitions. Second, for presentation purposes, we select one individual firm-level performance variable, total sales per employee, which is chosen because it provides the clearest measure of firm efficiency available from our data. Table 2 presents the full ML results for this two-equation system. Near the bottom of Table 2, one sees that the estimate for ρ is positive and a Chi-square test strongly suggests that the two error terms from (1) and (2) are in fact correlated. Thus, OLS estimation of (1) would lead to significant bias if the recovered estimate for α were used by itself to calculate the effect of firm clusters on

[7] A full set of results for the 104 systems is available from the authors. Tables 3 and 4 draws from all 104 systems to summarize the impact of clusters on various firm decisions, and illustrates how this impact is sensitive to the definition of a firm cluster.

sales per employee. In addition to the results for one particular system estimation, we summarize select results from all 104 systems in Table 3, which presents the 104 estimates of α.[8]

i. Cluster Formation

The first numeric column of Table 2 presents the ML results that reflect how firm characteristics and economic conditions influence the formation of a six-firm cluster of organic handlers. Nine of the 30 coefficient estimates (excluding the constant term) are found to differ significantly from zero. Four of the nine describe firm functions or specialization: Firms that function as a broker, and firms that have both production and handling functions, are less likely to be in a cluster. On the other hand, firms functioning as packers or shippers, and independent firms with only one facility are more likely to be in a cluster. Three other of the variables with significant impacts generally describe input availability: The number of small farms in a county has a positive impact on clustering, while a firm's priority for year-round supplies and a firm's total number of organic suppliers both have a negative impact on clustering. Finally, two demographic variables – population and the percent of the population with college degrees – both have a positive impact on clustering.[9]

Taken collectively, several of these results provide insight into the potential positive and negative tradeoffs from firm clustering that stem from competitive and cooperative behavior. Firms with many organic suppliers may feel disadvantaged in a cluster because of increased competitive pressures. On the other hand, independent one-

[8] Actually, Table 3 presents only 101 estimates of α because a joint ML estimation of the two-equation system fails to converge in three instances. While a two-step procedure is successful in recovering estimates of α, these results are omitted in Table 2 to preserve a more direct comparison of estimates.
[9] Though not presented in Table 2, similar results generally hold when the clustering definition, C_n, varies for n = 4, 5, 7, and 8.

facility firms may find advantages if clustering firms can increase scope economies through cooperation. Firms that produce as well as handle, however, may see less need for cooperation and therefore be less inclined to cluster.

ii. Impacts of Clusters on Firm Efficiency

The second numeric column of Table 2 shows how the cluster variable, C_6, and other factors impact firm efficiency as measured by sales per employee. First and foremost one should see that the coefficient on C_6 is negative and significant. By itself, this estimate would lead one to believe that firm clusters have a detrimental effect on firm efficiency. However, as equation (5) shows, and as we discuss below, the true estimate of the average effect of clusters must take into account the impact of sample selection bias.

Apart from the cluster variable, eight of the other 25 coefficient estimates (excluding the constant) are statistically significant. All else equal, firms with multiple locations are more efficient than the single-location firms, as are firms that self-identified themselves as large. Food manufacturers/processors and packer/shippers are less efficient than firms with other functions. And firms that experienced shortages of organic ingredients ended up being more efficient than firms that did not. Finally, firms located in more populated counties and in counties with higher nonfarm per capita income are also more efficient. Some of these results may suggest productivity gains from specialization or returns to scale. Others may suggest gains from a stronger labor pool.

iii. Impacts of Clustering on Additional Firm Decisions

Tables 3 and 4 present results from all 104 estimated ML systems showing (i) the coefficient estimates for α in each case, and (ii) the selection bias-adjusted average effect from clusters for the wider range of firm-level decision variables or firm performance measures. Table 3 shows that the estimated α is statistically significant in most models (i.e., in 57 out of 101 ML estimations that converged). It also shows that the sign and level of significance can vary across the two dimensions depicted in the table: variation in the dependent variable, and variation in the minimum number of firms used to define a cluster. An extreme example of this variation is found in the system with the dependent variable "Organic procurement – % local", which captures the percentage of total organic procurement that is done at a local level. For this case, Table 3 shows that the estimate for α is positive and significant when a three-firm cluster is used (C_3), but negative and significant when an eight- or nine-firm cluster (C_8 or C_9) is used. Looking at all 104 cases, however, Table 3 shows that when the estimate for α is statistically significant, the sign of α is generally stable.

Table 4 uses recovered estimates of α plus estimates for ρ and other recovered information to calculate the average cluster effect, accounting for selection bias, given by equation (5). The first row of Table 4 shows the average impact of clusters on total gross sales per employee. Note that for the C_6 column, Table 4 shows that a firm cluster (defined with a six firm minimum) leads to an average gain of $1.32 million in increased sales per employee. This positive value for a cluster's impact contrasts with the corresponding negative estimate of α reported in Tables 2 and 3. Following equation (5) for this example, a positive estimate for ρ helps overcome a negative estimate for α. This finding suggests that firms that are naturally more likely to benefit from clusters do in fact seek out clusters. A good example of this effect, mentioned in the discussion of Table 2, concerns firms that are independent with a single organic facility. These firms, apparently, choose to cluster, and in turn expect the cluster to provide efficiencies in scope that would otherwise be unavailable to an independent firm.

Table 4 furthermore shows that firm clusters have a positive (i.e., beneficial) impact on sales per employee. Clusters also have a positive impact on a firm's percentage of organic sales and organic procurement. Clustering positively impacts the percentage of sales and procurement in local markets, while negatively impacting the percent of sales and procurement in national markets. Perhaps the most noticeable impacts concern procurement decisions. The presence of a cluster increases the organic component of total procurement by as much as 37.2 percent (in column 10); it increases organic procurement made locally by as much as 23.4 percent (in column 6); and it decreases procurement from national markets by as much as 32.5 percent (in column 6).

iv. Sensitivity of Impacts to Cluster size

Table 4 also shows that the impacts from clusters are sensitive to the definitional size of a cluster. A good example of this sensitivity concerns total sales per employee. When the minimum number of firms used in the cluster definition is small (e.g., $n = 3$) the impact is relatively small ($0.17 million). However, this impact increases as n increases: For $n = 8$, the average effect of a cluster is an additional $1.44 million sales per employee. Other dependent variables show similar sensitivity. For low values of n, the impact of clusters on full-time employees is somewhat minimal (ranging between 9.0 fewer or 8.7 additional employees when n equals 3, 4, or 5), but the impact is more dramatic for higher levels of n. When $n = 10$, for example, firms in a cluster have on average 93.6 fewer employees. In several cases, the sign of the average impact changes as n changes: For example, firms in a C_5 cluster have $11.4 million more in total gross sales than non-clustered firms, on average. On the other hand, firms in a C_9 cluster have $13.7 million less in total gross sales than non-clustered firms, on average. In many instances, the cluster effect intensifies as the definitional number of firms in a cluster increases.

v. Robustness of Results and Model Specification

In addition to experimenting with variations of the cluster definition based on the minimum number of firms, we also investigated several other specification issues to see how robust our results were. First, instead of basing clusters on counties, we replaced county borders with the geographic boundaries that follow the first three digits of the U.S. zip codes. Second, instead of fixing the sample size for estimation at 316 observations, we allowed the number of observations to vary for the estimation of the 104 ML systems. Because each system's estimation relies on different dependent variables, and because not all questions in the USDA Economic Research Service's survey were answered with the same frequency (see Dimitri and Oberholtzer 2008) we allowed the econometric software to pick the maximum number of observations for each estimation. And third, we experimented with subsets of the $\boldsymbol{x}$ and $\boldsymbol{z}$ vectors to estimate the ML systems. More specifically, we removed some of the category (iv) dummy variables from equation (1) listed in Table 2, and also some elements of $\boldsymbol{x}$ that were potentially endogenous (such as a dummy variable for "Year-round availability priority"). In all of the model specifications described by these cases, we found almost no substantial differences in our estimation results described by Tables 3 and 4. These robustness checks, therefore, provide an increased sense of confidence.

Discussion and Conclusion

This paper expands the sparse literature that attempts to document the impacts of firm agglomeration (or firm clusters) on firm-level performance or firm-level decisions. We investigate the certified organic handler sector, a specialized component of the middle part of the farm-to-table marketing chain. To estimate the impacts from clusters, however, we have at least two preliminary tasks. First, in an attempt to explore how firm clusters might be defined, we allow the definition of a cluster to vary by the minimum number of similar firms present in a geographic area. Second, we draw on the treatment effects literature to estimate clusters' impacts only after accounting for possible endogeneity in the formation of clusters.

While our study confirms that endogeneity often is an issue that could bias results, our most important findings confirm that firm clusters have significant impacts in the organic handling sector. The exact measurement of clusters' impacts, however, depends on how a firm cluster is defined. For example, significant impacts on sales per employee range from an additional $0.17 million to $1.47 million, depending on whether a small or large number of firms is used as the minimum number to define a firm cluster.

Taken collectively, our results also help shed light on the tradeoffs between competition and cooperation in the organic handling sector. For this sector, our results suggest that cooperative forces may outweigh competitive forces. For example, one might be tempted to speculate that organic handlers in a county-based cluster of at least seven or eight similar organic firms may be forced out of local markets for their procurement because of intense competitive pressures. However, the opposite is true: clustered firms are more likely to procure locally. At least two reasons could explain this finding. First, as the firm agglomeration literature suggests, clustered firms attract organic production and create a strong local economy that can support a large cluster. Second, clustered firms in our data may be heterogeneous enough to create sufficient synergies in operations, thereby using other firms to create economies of scope. In other words, one possible explanation for the positive impact of clusters is that these organic handlers are middlemen who buy and sell from each other, thereby allowing more specialization to occur.

The above discussion notwithstanding, intense competition is still evident in our results, particularly when the definition of a firm cluster is based on a large number of firms. For example, firms in a large cluster may have, on average 75 to 95 fewer employees than other firms. It is interesting to note, however, that the same clustered firms (using a C_9 cluster definition, for example) see sales per employee increase by \$1.47 million while the total number of employees decreases by 74.9 employees. Here, the competition for labor may be so intense that firms have adjusted their operations to be less labor intensive and more output efficient.

References

Cainelli, G. 2008. "Spatial Agglomeration, Technological Innovations, and Firm Productivity: Evidence from Italian Business Districts." *Growth and Change* 39: 414-435.

Cameron, A.C., and P.K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press, New York.

Davis, D.E., and G.E. Schluter. 2005. "Labor-Force Heterogeneity as a Source of Agglomeration Economies in an Empirical Analysis of County-Level Determinants of Food Plant Entry." *Journal of Agricultural and Resource Economics* 30: 480-501.

Dimitri, C. and L. Oberholtzer. 2008. "The U.S. Organic Handling Sector in 2004: Baseline Findings of the Nationwide Survey of Organic Manufacturers, Processors, and Distributors," *Economic Information Bulletin* 36, U.S. Department of Agriculture, Economic Research Service.

Ellison, G. and E.L. Glaeser. 1997. "Geographic Concentration of in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy* 105: 889-927.

Ellison, G., E.L. Glaeser and W. Kerr. 2007. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." NBER Working Paper 13068.

Gabe, T. 2008. "Impact of Agglomerations on the Economy," Ch. 6 in Goetz, S.J., S. Deller and T. Harris, eds., *Targeting Regional Economic Development*, Routledge Publ., Abingdon, Oxon, UK, in press.

Gabe, T.M. 2004. "Industry Agglomeration and Investment in Rural Business." *Review of Agricultural Economics* 27: 89-103.

Gibbs, R.M. and G.A. Bernat. 1997. "Rural Industry Clusters Raise Local Earnings." *Rural Development Perspectives*, 12(3), 18-25.

Goetz, S.J. 1997. "State and County-Level Determinants of Food Manufacturing Establishment Growth: 1987-1993." *American Journal of Agricultural Economics*, 79 (August), 838-850.

Goetz, S.J., M. Shields and Q. Wang. 2008. "Identifying Food Industry Clusters: A Comparison of Analytical Tools," Ch. 15 in Goetz, S.J., S. Deller and T. Harris, eds., *Targeting Regional Economic Development*, Routledge Publ., Abingdon, Oxon, UK, in press.

Graham, D.J. and H.Y. Kim. 2008. "An Empirical Analytical Framework for Agglomeration Economies." *Annals of Regional Science* 42: 267-289.

Greene, W.H. 2008. Econometric Analysis, Sixth Edition. Pearson Prentice Hall: Upper Saddle River, New Jersey.

Porter, M.E. 1998. *On Competition*. Harvard Business Review, Boston, MA.

Porter, M.E. 2003. "The Economic Performance of Regions." *Regional Studies* 37, 549-578.

Roe, B., E.G. Irwin, and J.S. Sharp. 2002. "Pigs in Space: Modeling the Spatial Structure of Hog Production in Traditional and Nontraditional Production Regions." *American Journal of Agricultural Economics* 84: 259-278.

Table 1: Descriptive Statistics for Data Used in the Empirical Models

Variable	Description/ Units	Min.	Mean	Max.
Endogenous Cluster Variables: C_n				
C_3 (i)	0,1	0	0.560	1
C_4 (i)	0,1	0	0.472	1
C_5 (i)	0,1	0	0.411	1
C_6 (i)	0,1	0	0.332	1
C_7 (i)	0,1	0	0.313	1
C_8 (i)	0,1	0	0.275	1
C_9 (i)	0,1	0	0.247	1
C_{10} (i)	0,1	0	0.215	1
Endogenous Firm-level Variables: y				
Total gross sales per employee	$	500	557,682	20 mil.
Total gross sales	$	250,000	13.3 mil.	150 mil.
Total full-time equivalent employees	#	0.25	55	750
Percentage of organic procurement	%	0.05	43.887	100
Percentage of organic sales	%	0	40.961	100
Organic sales – % local (w/in 1 hour)	%	0	24.081	100
Organic sales – % regional (bordering states)	%	0	29.768	100
Organic sales – % national	%	0	38.675	100
Organic sales – % international	%	0	7.608	100
Organic procurement – % local	%	0	23.019	100
Organic procurement – % regional	%	0	29.816	100
Organic procurement – % national	%	0	24.312	100
Organic procurement – % international	%	0	22.821	100
*Exogeneous Variables (**x**) and category*				
Multiple locations* (iv)	0,1	0	0.250	1
Manufacturer/processor* (iv)	0,1	0	0.718	1
Wholesaler/distributor* (iv)	0,1	0	0.307	1
Broker* (iv)	0,1	0	0.076	1
Packer/shipper* (iv)	0,1	0	0.184	1
Years as a certified organic handler* (iv)	#	0	4.703	29
Years in business* (iv)	#	1	26.27	138
Certified organic producer and handler* (iii)	0,1	0	0.215	1
Animal food manufacturer* (iv)	0,1	0	0.092	1
Grain or oilseed milling* (iv)	0,1	0	0.168	1
Sugar or confectionery products* (iv)	0,1	0	0.057	1
Fruit or vegetable preserving*(iv)	0,1	0	0.146	1
Dairy product manufacturing* (iv)	0,1	0	0.104	1
Animal slaughtering or processing *(iv)	0,1	0	0.051	1
Bakery or tortilla manufacturing* (iv)	0,1	0	0.070	1
Beverage manufacturing* (iv)	0,1	0	0.095	1
Shortage of organic products (iii)	0,1	0	0.358	1
Priority of choosing local suppliers* (iii)	0,1	0	0.601	1
Self-identified facility size* (iv)	1=small, 2=med., 3=large	1	1.465	3
#of farms with size 10-49 acres* (iii)	#	0	513.40	2,928
# of farms with size 50 acres or more* (iii)	#	0	60.36	522

Table 1 (continued)

Variable	Description/ Units	Min.	Mean	Max.
*Exogeneous Variables (**x**) and category (cont'd)*				
Market value of land/building per acre* (ii)	$	136.00	5,731.54	48,159.00
Education - college* (ii)	%	8.2	23.48	54.6
Nonfarm Income Per Capita* (v)	$	15.60	31.06	73.99
Population* (ii)	#	2,160	731,380.9	9,880,732
*Exogeneous Variables (**z**) and category*				
Recruits existing organic suppliers (iii)	0,1	0	0.465	1
Year-round avail. a main priority (iii)	0,1	0	0.661	1
Using contracts for procurement (iii)	0,1	0	0.725	1
Independent with 1 cert. organic facility (iv)	0,1	0	0.737	1
% procured from spot market (iii)	%	0	29.241	100
Total # of certified organic suppliers (iii)	#	0	12.541	750
Number of complete observations = 316				

Notes:
* Exogenous variables in ***x*** marked with an * are also included in ***z***.
Categories for C_n, ***x*** and ***z***: (i) firm agglomeration variable, (ii) urban encroachment and population characteristic variables, (iii) input availability variables, (iv) firm productivity and specialization variables, (v) local economic variables,

Table 2: ML Results for Cluster Formation, C_6, and Sales Per Employee
(z-stats in parentheses**)**

	Dependent Variables	
	C_6	Total Gross Sales per Employee
Constant	-2.01** (-2.92)	-219,961 (-0.32)
Cluster (C_6)	---	$-2.22\text{x}10^{6}$** (-9.54)
Multiple locations	0.290 (1.08)	413,873* (1.61)
Manufacturer/processor	-*0.023 (-0.09)	-995,182*** (-3.49)
Wholesaler/distributor	0.239 (1.17)	331,225 (1.31)
Broker	-0.677* (-1.95)	-269,626 (1.31)
Packer/shipper	0.146 (0.60)	-562,032* (-1.87)
Years certified organic handler	-0.053 (-0.28)	-5,467 (-0.25)
Years in business	-0.004 (-1.11)	-3,552 (-0.79)
Certified organic producer and handler	- 0.382* (-1.71)	-361,526 (-1.32)
Animal food manufacturing	0.018 (0.06)	-138,824 (-0.33)
Grain or oilseed milling	0.298 (1.18)	549,710* (1.64)
Sugar or confectionery products	-0.662* (-1.71)	-535,636 (-1.10)
Fruit or vegetable preserving	0.428 (1.59)	316,830 (0.98)
Dairy product manufacturing	-0.214 (-0.77)	-408,393 (-1.13)
Animal slaughtering or processing	-0.102 (-0.24)	-446,429 (-0.90)
Bakery or tortilla manufacturing	-0.182 (-0.48)	-345,207 (-0.78)
Beverage manufacturing	-0.114 (-0.38)	-361,090 (-0.96)
% procured from spot market	0.001 (0.60)	---
Total # of organic suppliers	-0.0.29*** (-12.74)	---
Shortage of organic products	---	334,290* (1.92)

Table 2 (Continued)

	Dependent Variables	
	C_6	Total Gross Sales
Recruiting existing organic suppliers	-0.059 (-0.45)	---
Choosing local suppliers a priority	0.511** (2.62)	335,485 (1.37)
Year-round availability a main supplier priority	-0.039 (-0.27)	---
Using contracts for procurement	0.113 (0.81)	---
Independent with one certified organic facility	0.505* (1.94)	---
Self-Indentified Facility Size	0.023 0.15)	308,167* (1.77)
# of farms with size 10-49 acres	0.001*** (2.92)	312.0 (1.24)
# of farms 50 acres or more	0.000 (0.21)	520.3 (0.33)
Market value of land/building per acre	2.97x10^{6} (0.18)	1.44 (0.07)
Education - college	0.025** (2.13)	---
Population	6.65x10^{7}*** (3.96)	0.180*** (2.64)
Nonfarm income per capita	-0.003 (-0.13)	40,878** (2.38)
ρ	0.930	
LR test $\rho = 0$:	$\chi^2(1)$=65.50, Prob > $\chi^2(1)$ = 0.000	

Notes: *** = statistically significant at the 99 percent level; ** at the 95 percent level; * at the 90 percent level.

Table 3: Signs of ML Estimates for α, with Different Cluster Definitions

	Clusters, C_n, where n = 3 to 10							
Dependent Variable	n = 3	4	5	6	7	8	9	10
Total gross sales per employee ($ millions)	+	–***	–***	–***	–***	–***	–***	*nc*
Total gross sales ($ millions)	+***	+***	–	–	–	+	+***	–
Total full-time employees	+**	+*	+	+***	+***	+***	+***	+***
Percentage of organic sales	–**	–**	–*	–	–	–	–	–
Percentage of organic procurement	–***	–*	–**	–**	–***	–**	–***	–***
Organic sales - % local	–***	–***	–	+	+	+	–	–
Organic sales – % regional	+***	+**	+*	+	+	+	–	+
Organic sales – % national	*nc*	–	–	+	+	+	*nc*	*nc*
Organic sales – % international	–***	–***	–***	–***	–***	–**	+***	–
Organic procurement – % local	+***	+	–	–	–	–*	–**	–
Organic procurement – % regional	–**	–**	–**	–	–	–	–	–**
Organic procurement – % national	+***	+***	+***	+***	+***	+***	+***	+***
Organic procurement – % internat'l	–**	–*	–**	–	–	–	–	–

Notes:

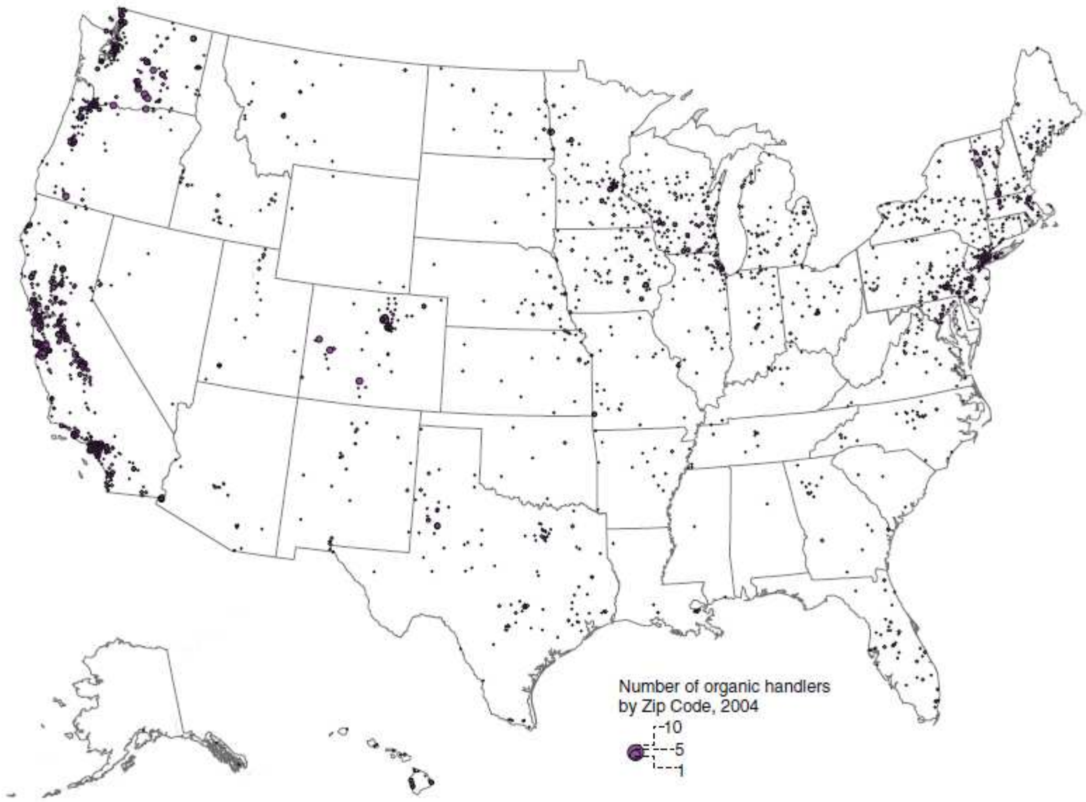
*** = α statistically significant at the 99 percent level; ** at the 95 percent level; * at the 90 percent level.
nc = two-equation ML estimation did not converge.

Table 4: Average Firm-Cluster Effect Accounting for Selection Bias, with Different Cluster Definitions

	Clusters, $\boldsymbol{C_n}$, where n = 3 to 10							
Dependent Variable	n = 3	4	5	6	7	8	9	10
Total gross sales per employee ($ millions)	0.17	0.43	0.98*	1.32*	1.38*	1.44*	1.47*	*nc*
Total gross sales ($ millions)	3.24*	0.05*	11.4*	9.47	11.1	3.88	-13.7*	3.63
Total full-time employees	8.7	-9.0	8.6	-53.7*	-50.4*	-79.0*	-74.9*	-93.6*
Percentage of organic sales	6.3*	9.7*	8.4	11.8	12.8	6.1	11.6	17.5
Percentage of organic procurement	8.3*	11.2*	8.9*	18.8*	24.5*	20.8*	27.2*	37.2*
Organic sales - % local	9.0*	10.1*	9.3	8.7	7.7	5.6	7.9	11.3
Organic sales – % regional	-7.7*	-10.1*	-9.1*	0.4	-3.7	-2.2	-3.6	0.3
Organic sales – % national	*nc*	5.2	-0.5	-14.1	-7.4	-6.3	*nc*	*nc*
Organic sales – % international	3.9*	4.9*	6.5*	3.1*	4.3*	2.9	-19.2*	-1.7
Organic procurement – % local	2.8*	5.9	17.3	23.4*	23.3*	22.5*	20.6*	16.9
Organic procurement – % regional	2.9*	1.0*	1.4	-0.7	-2.5	-9.9	5.0	12.8
Organic procurement – % national	-20.9*	-23.5*	-26.8*	-32.5*	-30.4*	-28.3*	-24.3*	-26.1*
Organic procurement – % internat'l	8.4*	12.6*	16.0*	15.4*	15.0*	18.0	5.3	9.6

Notes:
*= estimates of ρ are statistically significant at the 90 percent level or better.
nc = two-equation ML estimation did not converge.

Figure 1: The Geographic Dispersion of Certified Organic Handlers (Dimitri and Oberholtzer, 2008)



Source: Economic Research Service, USDA.