AGRO-INDUSTRY LOCATION:
THEORY AND TEST

Draft of work in progress

by
Maureen Kilkenny and Scott Coleman[1]

[1] Kilkenny is an Associate Professor in the Dept. of Resource Economics, University of Nevada, Reno. Scott Coleman recently earned an MS in Economics at Iowa State University. The empirical work in this paper was conducted by Scott Coleman in partial fulfillment of his graduate degree requirements.

The research was supported initially by co-operative agreement 43-3AEM-8-80122 between Kilkenny while she was at Iowa State University and the Economic Research Service of the USDA, and later by an assistantship provided to Scott Coleman by Prof. Dermot Hayes at Iowa State University. The views expressed herein are solely those of the authors. They thank Gerry Schluter, Jacques Thisse, and Wolfgang Kliemann. Please do not quote without permission from kilkenny@unr.edu [775] 784-6785.

AGRO-INDUSTRY LOCATION: THEORY AND TEST

ABSTRACT: This paper poses and tests a new model of the location of plants that process ubiquitously supplied 'von Thünen' inputs such as raw farm output. The new model rationalizes why agro-industry is city-located but resource-oriented. It demonstrates the significance of relative transport costs in agro-industry location. Furthermore, we present a unprecedented direct test/empirical confirmation of establishment location theory.

We develop this agro-industrial plant location model in two stages. In the macro-spatial stage, we show why an initial agro-industrial plant chooses a site that becomes a city, and how that initial site can establish a market center and von Thünen rent gradient. In the micro-spatial stage we show that even though subsequent plants can choose the extent of their input market areas as well as their site, costs are convex (not concave) in distance. Thus, in contrast with canonical location theory, intermediate sites can be optimal for agro-industrial plants, even in the absence of uncertainty.

This new testable model rationalizes the observed variety in spatial configurations of U.S. agro-industry. It justifies the use of normative math programming models of food and kindred processing plant location widely used by agricultural economists. The model is tested using data on transport cost rates, crop yields, acreages, and existing agro-industrial plant locations. The null hypothesis is strongly rejected for all agro-industries modeled. In the U.S. these types of plants systematically display location patterns that are concentrated, co-located, or dispersed, depending on relative transport costs in a predictable way.

1. INTRODUCTION

The motivation for this paper is to rationalize why agro-industrial plants that compete with each other for inputs neither fully disperse nor concentrate spatially. For example, many agro-industrial plants in USA appear to have overlapping input market areas, resulting in higher prices paid for inputs in the areas where their input markets overlap (Figure 1). Why wouldn't they more fully disperse? Location theory is incomplete, and the empirical evidence (e.g., Goetz, 1997; Kim, 1999) about agro-industrial location appears paradoxical. In this paper we attempt to fill the gap in location theory in a way that reconciles the paradoxes.

Even in the latest theoretical papers focusing on agro-industrial location (c.f., Hsu, 1997) the conclusions are that profit-maximizing agro-industrial plants optimally locate either at the product market center or as far away from it as possible. At a macro spatial level the implications are that agro-industry would be found either in metropolitan or rural, but not urban counties. But the U.S. data do not support this implication. There is disproportionately more agro-industry in urban than metro counties, and only

3% of U.S. agro-industry establishments are in rural counties (Kilkenny and Schluter, 2001). U.S. agro-industry appears to be both “materials-oriented” *and* market-located.

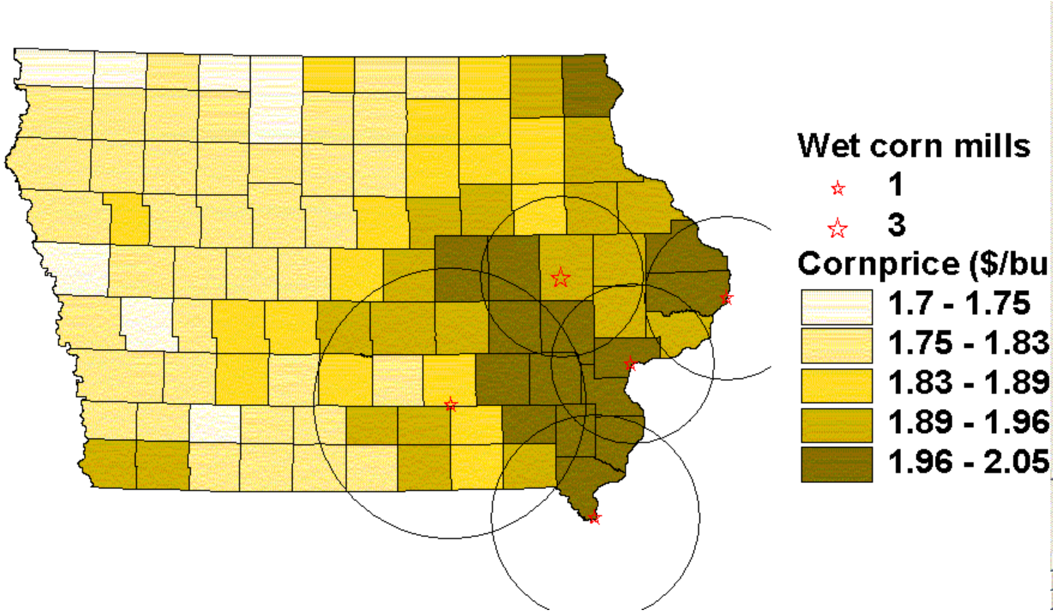


Figure 1. Iowa’s 7 Wet Corn mills and county-wide farm-gate prices of corn (2000)

Although none of the theoretical papers pose empirically testable hypotheses (much less test them), other empirical researchers have concluded that U.S. agro-industry is “market-oriented” (e.g., Goetz, 1997) because market size has been significantly positive in county-level regression analyses of the counts of food and kindred processing establishments. Goetz used population as a proxy for both the size of the market and localization or agglomeration economies. Larger populations were found to be associated positively with agro-industry growth overall. Goetz also found that counties with higher historical food processing concentrations lost establishments rather than gained them, which he interpreted as evidence of dynamic localization diseconomies of scale.

Our alternative interpretation is that new plants may be repelled from the sites of existing plants with whom they might compete for inputs from the surrounding input-supplying region. While other analyses of SMSA-level food processing sector production functions have found what may be positive localization economies, those estimated effects have been shown to be insignificant when the extent of nearby agricultural input production is taken into account (Sveikauskas, Gowdy, and Funk, 1988).

In direct contrast, other researchers have characterized U.S. agro-industry as dispersed and "input-oriented." The locational Gini coefficient for the food and kindred processing sector at the county level of observation is 0.15 (Barkley and Henry, 1998). This is significantly lower than the 0.30 average for all U.S. sectors, indicating that agro-industry is one of the more widely dispersed. And in a state-level empirical analysis, Kim (1999) showed that food and kindred processing is resource-oriented. The observed macro spatial dispersion across related U.S. farm regions parallels the dispersion of the land resources used intensively in the production of the agricultural inputs used by each agro-industry. Finally, agro-industry also appears to disperse to avoid competition for inputs (Sexton, 2000).

The inconsistencies between these alternative strands of empirical or descriptive evidence and the hypotheses proposed to date by location theorists have prompted some agricultural economists to state that "very little is known about the determinants of firm location," Goetz (1997) page 838.

Indeed, not much is known about the determinants of agro-industrial establishment location in theory. Hsu's 1997 and 1999 articles are the only analytical treatments of the problem of establishment location with respect to ubiquitously supplied inputs of the 'von Thünen' type: whose prices decline with distance from the market. Hsu's models are particularly well-suited to analyzing agro-industrial location on small islands such as Taiwan. He assumes that the entire input market area of an agro-industrial plant is predetermined and covers all possible sites for a subsequent plant within the input market area of any initial plant. But these assumptions do not apply to North America, Europe, China, Australia, or African continental cases where distances are vast and no agro-industrial establishment is so large that it's input market area covers the whole region. There does not appear to be a model of agro-industrial plant location relevant to continental areas (Kilkenny & Thisse, 1999).

Therefore, in this paper we present a model of agro-industrial location suitable for continental areas. Furthermore we present a simple, preliminary test of the model against the data. A theory is a hypothesis that has not been rejected by the data. Our new model generates implications consistent with the observed preponderance of food and kindred sector manufacturing establishments in U.S. urban and metro, rather than rural, counties.

To our knowledge this is the first direct test of establishment location theory. Other empirical models of establishment location are "hedonic," in which existing plants are modeled to choose a site because of the characteristics of the site. We know of no other empirically tested model derived directly from the first order conditions of an establishment location choice problem. In the canonical establishment location problem (see Hurter and Martinich, 1986) the key explanatory arguments are the trade-offs between fixed costs and transport cost rates and input:output relations, known among location theorists as *ideal weights*. The canonical model is the basis for countless normative and simulation models. But it has not been testable until now.

2. Agro-industrial Plant Location Theory

We develop the new location theory for agro-industrial plants in two stages: macro-spatial then micro-spatial. First we establish the optimal location of an initial plant relative to the growing region. Not surprisingly we show that the initial plant's optimal location is central to the input supply region regardless of whether output or input transport is cheaper. Furthermore, the extent of the input market area depends, as in canonical location theory, on fixed costs, value-added, and transport costs. Then, because places where the first plant opens becomes a point mass of population, that first site becomes an optimal location for subsequent market-oriented establishments. Subsequent activity locates near the bulk of the customers to maximize profits by minimizing transport costs.

If agglomeration economies of scale also exist, they would reinforce the concentration of establishments in *market* sites, i.e., in cities. But we must emphasize that agglomeration economies of scale are not necessary to rationalize the metro locations of agro-industrial plants. Transport cost savings alone justify it. In either case, the first plant establishes an urban core and it also sets up the farm-gate price gradient for farm output that declines with distance.

Then we consider micro-spatial location. We show that the optimal locations of subsequent agro-industrial plants are not necessarily where the first plant is, nor are they as far away as possible. That is, we find that profits can be strictly concave in distance. For location theorists this is an unexpected result. It implies that the optimal site of subsequent plants may be at the initial site location, in-between, or in the otherwise undeveloped hinterland, depending on relative transport costs. This shows that the appearance of attraction (or repulsion) between plants does not necessarily depend on positive agglomeration externalities or violations of perfect competition assumptions about input prices.

To derive our new model, we follow canonical location theory and assume that the initial plant pays a uniform delivered price for inputs. Like Hsu (1997) we assume that farmers are uniformly dispersed across space. In consequence, farm product prices and thus agricultural land values decline with distance from the center. But in contrast with Hsu (1997), because we assume that plant location and the input market area are both choice variables, we identify conditions under which subsequent plant market areas center on the output destination (which we call "concentration"), overlap prior plants' input market areas ("co-location"), or are non-overlapping ("dispersion").

3. MACRO SPATIAL LOCATION: AGRO-INDUSTRY IS INPUT-ORIENTED AND CITY-LOCATED

To site an initial plant that processes a ubiquitously-supplied input, imagine an otherwise featureless plain across which N farmer-consumers are uniformly distributed. Without loss of generality, the dimensionality of the problem can be reduced from planar space to linear space (Lederer and Hurter, 1986). Index locations along the linear space by (*m*) as in mile, (*s*) as in site, and (*f*) as in farm. Farmers produce a raw product, "corn," that has opportunity cost ($\underline{P}$) which we assume covers the cost of production. Farmers can transport their corn to a processor at the input transport cost rate (*t*) per unit product and distance. They'll grow and market corn if their net return at the market price, P_c, is at least $\underline{P}$, i.e., if $P_c - t|f\text{-}s| \geq \underline{P}$. Then, as customers of the processed product, each farmer demands one unit at the full or delivered price, P_D.

An agro-industrial firm will establish a plant, buy, process, and transport output to customers on farms (&/or at the plant site) if total revenues cover total costs. Costs include the sunk cost of plant capital (K), average (constant marginal) processing costs per unit (c); and, by proper choice of units, one unit corn input per unit output. Total production costs are $K + cN + P_cN$. Given an output transport cost rate (T), the revenue from a customer at location (m) is $P_M = P_D - T|s-m|$. Therefore revenues at the plant site (s) are $P_DN - \int_0^N T\,|s-m|\,dm$, and profits are:

(1) $$P_DN - K - cN - P_cN - \int_0^N T\,|s-m|\,dm$$

The first four terms are independent of the plant location at site s. Only the last term, total transport costs, varies with respect to the site. This highlights the fact that the site that maximizes profit for the <u>initial</u> agro-industrial plant is the site that minimizes the costs of transporting output to an entirely rural population.

The optimal site will be one of the three types: on one side (s=0), the other side (s=N), or somewhere in the middle of the featureless plain. Evaluating the integral solves for total output transport costs with respect to any site, s:

(2) $$\int_0^N T|s-m|\,dm \;=\; T\left[\int_0^S (s-m)\,dm + \int_S^N (m-s)\,dm\right] \;=\; T\left[s^2 - sN + \frac{N^2}{2}\right].$$

The first- and second-order conditions for a minimum show that $s^* = N/2$. The optimal location of the first single plant is in the center of its input (or output market area, since input suppliers are by assumption also the output demanders.) Note: the value function shows that total transport costs borne directly by the initial plant are $\frac{1}{4}TN^2$.

The optimality of the site central to the growing region is independent of both transport cost rates. But the extent of the input market area depends on both. The *extensive margin* or radius of the input market area around the initial plant, d*, is such that $d^* \leq (P_c - \underline{P})/t$, where, P_c, the price paid for corn is limited by the price customers will pay for the output (P_D), over average production costs and transport costs per unit output.[2] Average output transport costs (following Beckmann and Thisse, 1986) per unit are $\frac{1}{4}TN$, so the optimal extent of the input market area, d*, is:

(3) $$d^* \leq (P_D - K/N - \tfrac{1}{4}TN - c - \underline{P})/t.$$

[2] A process that requires γ units of input per unit output, uses $\Sigma_m q(m) = \Sigma_f q(f)\gamma$ inputs, and has an input transport cost rate of τ total transport costs are: $\Sigma_m q(m)\cdot T\cdot d|s-m| + \Sigma_f q(f)\cdot\gamma\cdot\tau\cdot d|f-s|$. Where $t=\gamma\cdot\tau$, input transport cost rates (per unit output) are lower for lower input:output ratios.

Equation 3 makes explicit many relevant spatial relationships. The size of the farming or rural region patronized by an agro-industrial plant is larger (i) the higher is the price people are willing to pay for the agro-industrial output, (ii) the larger is the consuming population, (iii) the lower are both input and output transport cost rates, (iv) the lower are the costs of production, or (v) the lower is the opportunity cost of corn (or the higher is farm productivity). Also, (vi), along the isoprofit curve the trade-off between fixed costs, K, and transport costs is dK/dT(or *t*) = $-\frac{1}{4}N^2$. These relationships are empirically verifiable.

We expect there to be fewer, larger plants, with larger market areas in developed or modern regions with good transportation infrastructure, or, for the types of products where transport cost rates are low (e.g., corn). We also expect to find smaller, more numerous plants in regions or for processes where transport is costly. At extremely high transport cost rates, the processing is optimally done on mobile platforms right in the field (e.g., vegetables). Equivalently, where transport cost rates are very high, farm production is likely to be solely for household subsistence (e.g., LDCs).

Also note that taking (K*,N*) as the optimal size of an agro-industrial establishment under (P_D, $\underline{P}$, c, *t*, and T) demand, technology, and transport cost conditions, a region with a population or market size of Z can support no more than Z/N* plants. This does <u>not</u> imply that Z/N* is the number of plants such that every farm everywhere is in at least one plant's input market area. There is no implication that every potential input supply area will have a local plant. That depends on the level of total demand relative to optimal plant size. When agro-industrial input demand is less than farm production capacity, not all potential sites will be occupied. Thus, even industries that are dispersed (relative to point-source industries) may appear to cluster where the markets are, leaving large areas of farm production to its next best alternative use, such as for feed. This would be particularly true when the consuming population is not uniformly distributed across space.

Furthermore, there is no reason to expect the consuming population to be uniformly dispersed across the farmlands. Even if it were dispersed initially, the establishment of industry requires that some farmers become workers. These workers move to the industry's site to minimize their own transport and commute costs. Thus, the opening of the agro-industrial establishment gives rise to a point mass of a consuming population in the center of the farm/rural region. In this way, the location (and productivity) of agriculture determines the locations (and sizes) of cities. Geographic historians have documented that indeed this has been true throughout history and around the world (Bairoch 1988; Mumford 1961).

The relatively dense area where industrial plants initially open may be classified as urban. Notice also that the initial agro-industrial site minimizes the cost of shipping to the urban consuming population (trivially true). And, the urban site for a first plant is optimal even if only workers, or only farmers, purchase the processed output. The central site's optimality is further reinforced if it subsequently becomes a transshipment point. It also remains optimal regardless of who is responsible for farm input transport; relative transport cost rates, or input:output rates. In sum, because agro-industry historically or initially locates to minimize the cost of obtaining inputs from a supply area, large agro-industry plants are likely to be in urban or metro counties even though they may be "input- or materials-oriented."

4. MICRO-SPATIAL AGRO-INDUSTRY PLANT LOCATION

Next, given the locations of farming regions and cities, consider the location choices of subsequent plants that (for simplicity) operate at the same scale, same technology, same output price, and the same transport cost rates as the initial agro-industry plants. The market for the processed output is no longer simply dispersed, it has a point mass of consuming workers at the center in addition to the farmers still dispersed across the hinterland. Farmers closer to the market can earn higher returns net of their delivery transport costs. This differentiates land values at different distances from the center (c.f. von Thünen).

Assume for simplicity that average farm production costs equal opportunity costs $\underline{P}$. Then farmer profits are $P_c - \underline{P} - t|f - N/2| - R_f$. For computational clarity, normalize distances by N/2 so that the site of the plant in the center is at 0. Competition and free entry drives farm profits to zero, and land rents at each site to $R_f = P_c - \underline{P} - tf$. Land rents decline from the city at the constant rate *t*, as long as cropping intensity is exogenous. Land rents decline at a decreasing rate otherwise. Furthermore, the price a farmer within the market area d* must receive for his corn in order to sell to anyone else must at least be as high as the opportunity cost, $\underline{P}$, plus rent (P_c- tf). Farmers earn no rent growing corn beyond the extensive margin; that is, beyond the d* such that P_c - $\underline{P}$- td* = 0. For this reason Hsu (1997) called the inputs used by agro-industrial plants "von Thünen" inputs.

Under similar, but not entirely the same assumptions, Hsu proved that a price-taking agro-industrial firm's profit function is decreasing and non-concave (convex) in distance (Hsu's Proposition 2). This is the condition for *end-point optimality* common to canonical location theory. Given the relative density of the output market, the obvious conclusion is that under Hsu's assumptions, all agro-industrial plants would locate at the market center.

The canonical *end-point optimality* or *exclusion theorem* is that only input locations, cross-roads, or market locations are optimal plant sites because the expected utility of profits is convex with respect to distance (e.g., Gruver and Xu, 1994; see Kilkenny & Thisse (1999) for a survey). Intermediate locations can be *excluded* from consideration. The property applies to a wide variety of location problems where input and output markets are at multiple discrete locations across space and (1) transport costs are positive, concave in distance, and all born by the firm in question; (2) the production technology is homothetic; and (3) the decision makers are risk-neutral (Hurter and Martinich, 1989).[3]

There is a long tradition of agro-industry location analysis that has been based explicitly or implicitly on the exclusion property: King and Logan, 1964; Takayama and Judge, 1964; Fuller, Randolph and Klingman, 1976; von Oppen and Scott, 1976; and Dunham, Sexton, and Song, 1995. Farm processing plant location problems have been posed as math programming problems where the objective is to determine the optimal scales, numbers, and locations of processing plants among given set of material source and market sink locations on a network. The problem has also been posed as a social planner's problem where the objective is to maximize producer plus consumer surplus (Takayama and Judge, 1964). In some applications, locations are entire regions rather than network nodes, and the distances between regions are measured between their geographic centers (e.g., Apland and Andersson, 1996). In sum, all

[3] The only conditions under which intermediate locations are not excluded are when decisive agents are risk averse and either demand prices at different market destinations and/or input prices from alternative input source locations vary stochastically. Under risk aversion and price uncertainty an intermediate location between inputs and markets is effectively a hedge.

extant models of agro-industrial location are either normative simulation models or hedonic descriptive models (Henderson and McNamara, 1997; Goetz, 1997). We know of no micro-foundations location model, neither for agro-industry nor point-source or point-sink location problems, that has been posed in a way that is amendable for statistical estimation or empirical testing.

The exclusion property abstracts from variations in land rent over distance. When the systematic variation in land rents a là von Thünen is taken into account, it should not be surprising that the exclusion property does not apply. The question is why the exclusion property *did* apply in Hsu's (1997) treatment of the agro-industrial plant location problem. The reasons are that Hsu did not include dispersed farmers as demanders of processed product, and he assumed that the extensive margin of all agro-industrial plants were co-terminous and exogenous. Hsu's firm chooses a site at $x \subset [0, d^*]$ and pays $P_c - tx + t(f-x) = P_c - tf$ (i.e., opportunity cost plus rent) to the farmers for their corn; plus wages and travel costs per unit of labor that resides in the city, plus the cost of transporting output back to the city. If the plant does not locate at the city, it bears the costs of remunerating commuting labor plus the cost of shipping output back to the city. Under Hsu's assumptions, *costs are concave* in distance from the city, profits are convex, the exclusion property applies, and endpoints are the only possible optimal sites.

Our assumptions *will* lead to the rejection of the exclusion property.

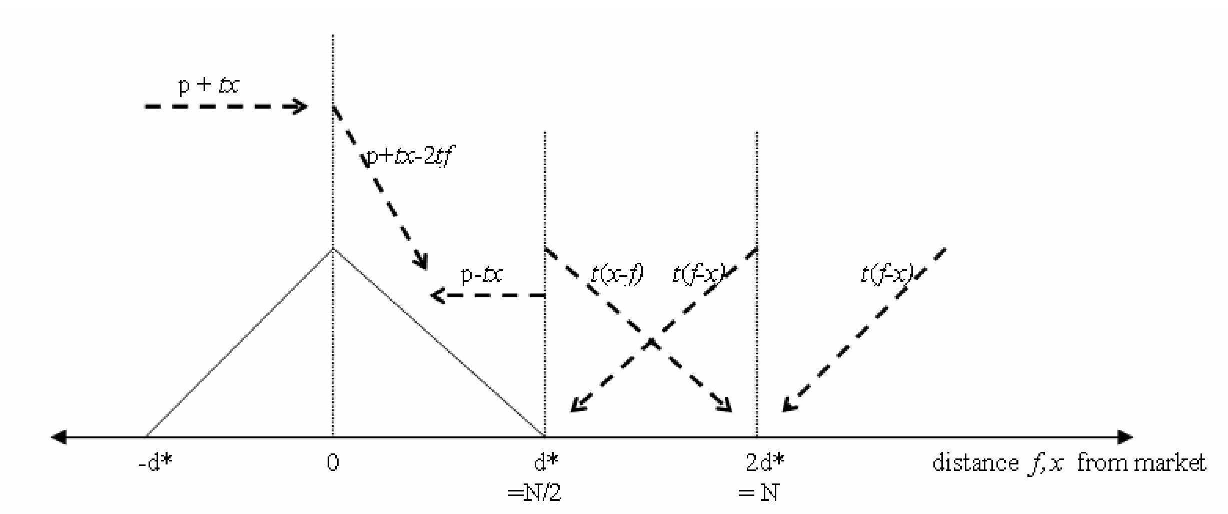
We allow a plant to locate on the same site (concentrate), within the original input market area (co-locate), or completely outside of it (disperse.) We also endogenize the boundaries of the subsequent plant's input market area. Under our assumptions costs are strictly *convex* in distance. The exclusion principle does not hold. Thus, city, rural, and intermediate locations can all be optimal sites for subsequent agro-industrial plants.

Formally, the subsequent plant chooses its input market area, and its site within it, over all possible sites inside and outside the initial plant's extensive margin; to maximize profits. The profit of a subsequent plant that operates at scale N and transports all output to the center from site x within its optimally chosen input market area $(x-a, x+N-a)$, is:

(4) $\quad \mathrm{P_D \cdot N - K - c \cdot N - N \cdot T \cdot}x - \int_{\mathrm{x\text{-}a}}^{\mathrm{x}+N-a} \max(\mathrm{P_c} - tf, \underline{\mathrm{P}}) + t|f-x| \quad df$

While the cost of output transport clearly increases with distance from the city, the number of farmers to whom the plant must pay the rent-ridden corn price decreases with that distance. Another useful normalization helps simplify notation and summarizes the rent: $\mathrm{p} = \mathrm{P_c} - \underline{\mathrm{P}}$. The economic geography is symmetric around the central city at 0, so we can focus only on one side of the market center. Due to the "von Thünen" nature of the inputs, the delivered cost of corn to a subsequent plant depends on the direction from the input origin (farm) to the destination (plant site), as shown in Figure 2.

Figure 2. Input price schedules with respect to distance



As indicated in Figure 2, a subsequent agro-industrial plant's site-related cost functions will differ for sites within either of the three possible domains: (i) between distances of 0 and N/2 from the market in which the subsequent plant and some of the input-supplying farmers are inside the initial plant's input market area, (ii) between N/2 and N in which the subsequent plant is outside, but some of the input-supplying farmers are inside the initial plant's input supply area, (iii) farther than N away from the market, in which both the plant and all input supplying farmers are outside the extensive margin of the initial plant's input supply region. Formally, these three types of site costs are:

(i) $\mathrm{N \cdot T \cdot}x + \int_{\mathrm{x\text{-}a}}^{x} \mathrm{p} + tx - 2tf \ df \ + \int_{\mathrm{x}}^{d} \mathrm{p} - tx \ df \ + \int_{\mathrm{d}}^{\mathrm{x}+N-a} t(f-x) \ df$

(ii) $\mathrm{N \cdot T \cdot}x + \int_{\mathrm{x\text{-}a}}^{d} \mathrm{p} + tx - 2tf \ df \ + \int_{\mathrm{d}}^{x} t(x-f) \ df \ + \int_{\mathrm{x}}^{\mathrm{x}+N-a} t(f-x) \ df$

(iii) $\mathrm{N}\cdot\mathrm{T}\cdot x + \int_{\mathrm{x\text{-}a}}^{x} t(x-f)\ df \quad + \int_{\mathrm{x}}^{\mathrm{x}+N-a} t(f-x)\ df$

Due to the monotonicity of the cost of delivered corn with respect to distance away from the plant in either direction, under general assumptions about transport costs, the minimum cost (maximum profit) input market area is defined by the extensive margins (x-a, x+N-a) where x is the site and a is a choice variable defining the patronized input supply area, chosen such that the gross price of corn at x-a is no more or less (i.e., equals) the gross price of corn at x+N-a. This '*patronize farmers in the cheaper direction'* principle implies that for plants located in either of the first two subsets of the domain, cases (i) and (ii), $a = \frac{1}{3}(x + \mathrm{p}/t)$. In the third case, the plant will cite itself in the middle of a wholly new input supply region: (iii) a=N/2.

Evaluating the integrals with endogenous endpoints for (i), (ii), and (iii) provides the plant's site-related costs in each of the three domains. These costs are:

5(i) $\mathrm{NT}x + a\mathrm{p} + a^2 t + (\mathrm{N}-a)^2 - atx + (\mathrm{d}-x)(\mathrm{p}-tx)$

5(ii) $\mathrm{NT}x + a(\mathrm{p} + at - tx) + \mathrm{p}(\mathrm{d}-x) + t[(\mathrm{N}-a)^2 - \mathrm{d}^2 + \mathrm{x}^2]/2$

5(iii) $\mathrm{NT}x + \frac{1}{4}t\mathrm{N}^2$

The derivatives of the cost minimization problem's first order conditions with respect to distance in the first two domains are strictly positive, indicating strict convexity of cost with respect to distance. The derivatives of the FOC with respect to distance in the third domain is zero. Thus, in all cases, we can conclude that costs (profits) are convex (concave) in distance.

The first-order conditions within the first domain implies that $\mathrm{x}^* = \mathrm{p} + at - \mathrm{NT}$. Given $a = \frac{1}{3}(x + \mathrm{p}/t)$ and $\mathrm{N} = 2\mathrm{p}/t = 2\mathrm{d}^*$, this implies that $\mathrm{x}^* = [2 - 3(\mathrm{T}/t)]\mathrm{d}^*$. Furthermore, $x^* = 0$ (where 0 is the site of the initial plant) when $\frac{\mathrm{T}}{t} \geq \frac{2}{3}$. The implication is that when output/input transport cost ratios are relatively high, inputs:output are low (to this point we have been assuming unitary input:output ratios), technology is relatively efficient, or there are multiple outputs; the 'ideal weight' of output is relatively high, and the optimal distance of the subsequent agro-industrial plant from the market center is low. This implication is consistent with the rule of thumb provided by canonical location theory.

The first-order conditions within the second domain (case (ii)) imply that $x^* = [14 - 27(T/t)]\ d^*$, which indicates that at relative transport costs more than ½, (14/27, or 22% lower than the lower bound on the first case)), the subsequent establishment's optimal site is outside of the initial input market area, but the optimal input market areas overlap. We call this co-location or polycentricity. The input market of a subsequent plant located right at d* optimally extends from ⅔d* on the left to 2⅔d* on the right.

The first-order conditions in the third domain imply that at even lower relative output transport cost rates, $\frac{T}{t} \le \frac{12}{27}$, the subsequent establishment will disperse. It will choose a site central to its own input market so that it does not compete for inputs at all. This is a site at least 2d* from the market center.

Given all the abstractions we have made, we avoid a strict or cardinal interpretation of these first-order conditions. In ordinal terms the principles are nevertheless clear. The envelopes of these site-specific *cost functions* are illustrated at various relative transport cost rates in **Figure 3**. The graphic illustration shows that costs are strictly convex in distance. Furthermore, under some relative transport cost rates, intermediate locations minimize costs (maximize profits).

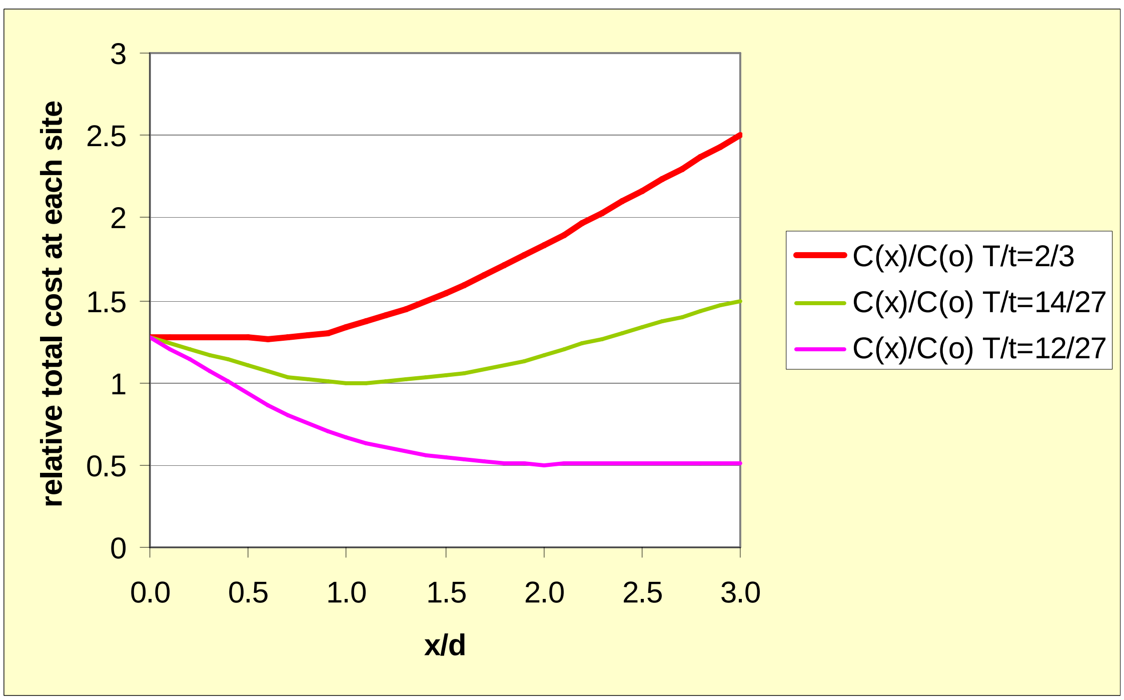


Figure 3. Total COSTS (profits) are strictly CONVEX (concave) in distance

5. EMPIRICAL TEST

This model predicts that if input transport cost rates, mill prices, and the input:output ratio are sufficiently high, agro-industrial plants in the sector will disperse. Taken together, those three factors comprise what locations theorists call the *ideal weight*. The other key explanatory variable is the radius of the input market area around the initial plant, d_e. We label the compound explanatory variable w_{ie} :

(6) $$w_{ie} = \frac{T_i}{t_i q_i} d_e^*$$

Our model implies that the revealed preferred distance x_{ie} , between the site of an agroindustrial plant establishment *e* in agroindustry *i* relative to its neighboring plant depends negatively on w_{ie} :

(7) $\boldsymbol{X} = \boldsymbol{\alpha Z} + \boldsymbol{\beta W} + \boldsymbol{\epsilon}$,

And the null hypothesis to be tested is that the relative concentration, co-location, or dispersion of plants is independent of the ideal weights.

The second innovation that renders our model testable is how we define and measure the variables. While output transport costs rates T_i, input transport cost rates t_i, and the input-output ratios q_i differ only across sectors, the distances between nearest-neighbor plants, x_{ie}, and the market radii d_e* of the two plants being observed, vary across all observations. This variation in our dependent and explanatory variables allows us to test the model in a way that no other specification has allowed.

In particular, as per Figure 4 below; our dependent variable is a discrete measure of whether a subsequent plant concentrates, co-locates, or disperses relative to an initial plant.[4] It takes just three categorical values corresponding to the three domains identified in Figure 2. First, $x_{ie} = 0$ if a plant and it's nearest neighbor are concentrated, that is, located within the same county; (ii) $x_{ie} = 1$ if the plants are co-located: not in the same county but having overlapping input market areas (so that the distance between the two plants in the pair is less than the sum of the extensive margins of the two plants); and

[4] Both the continuous and the discrete version of the dependent variable provided the same qualitative results, but the significance of the coefficient on ***W*** in the discrete version is much higher.

(iii) x_{ie} =2 if they are dispersed, i.e., if the actual distance between the two plants is greater than the sum of the extensive margins of the two plants in the pair.

(i) Concentrated (x_{ie} =0) (ii) Co-located (x_{ie} =1) (iii) Dispersed (x_{ie} =2)

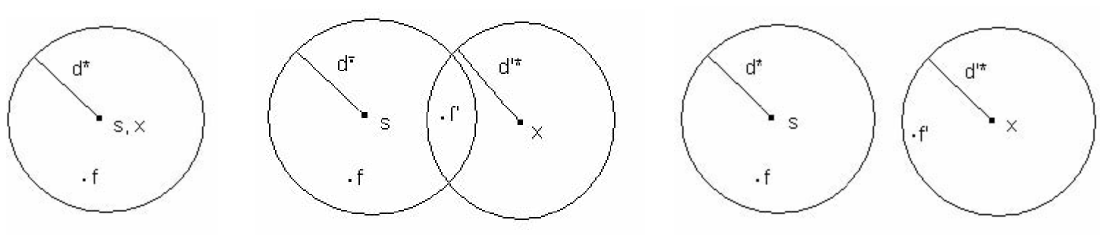


Figure 4. The dependent variable x_{ie}

If output *ideal weights* in an agroindustrial sector are higher, we hypothesize that plants in the sector will concentrate in market centers and the agro-industry would be theoretically correctly classified as "demand-", "output-" or "market-oriented", that is, x_{ie} will be smallest. It is also possible that at intermediate relative *ideal weights,* plants may co-locate or locate "somewhere in-between" in violation of the typical *exclusion property*, according to our new hypotheses above. There is no canonical classification for this type of industry, although "footloose" may be a reasonable proxy (Conner & Scheik, 1997). Finally, if the output ideal weight is low relative to input ideal weight, the plants are expected to be farthest apart; to have completely non-overlapping input market areas. In that case the agro-industry would be correctly classified as "supply-" "input-" or "materials-oriented."

To measure q_i, T_i, t_i, and x_{ie}, we obtained input-output, transport cost rate, and county-level data on the locations of all establishments with employees in seven agro-industries at the NAICS six-digit level of disaggregation. The industries studied are Dog and Cat Food Manufacturing, Flour Milling, Malt Manufacturing, Wet Corn Milling, and Soybean Processing, Sugar Manufacturing, and Fluid Milk Processing. These seven were chosen because they each process mainly one ubiquitous input (corn, wheat, barley, corn, soybeans, sugar beets, and milk; respectively), the data was relatively more complete (not suppressed), and they correspond to the range of "supply-oriented," "footloose," and "demand-oriented" agro-industries as classified by Connor and Schiek (1997).

The input-output data (for q) was provided by our colleague Gerry Schluter at ERS-USDA. Transport cost information (T and t) were obtained from internet web sites. We tediously collected at least nine quotes for overland truck freight transport for each input commodity and output product of interest for city origins and destinations about 500, 1,000, and 4,000 miles apart (three at each distance, different times of the year) in the continental USA. We regressed the quoted total transport cost per ton on the number of miles shipped on each route we quoted. Our estimated transport cost rates per ton per mile are the coefficients on miles shipped in each regression equation. We fit one equation for each input and output to estimate fourteen transport cost rates.

Then we created a data base of plant locations where the units of observation are plant-pairs, from data in which the units of observation are counties. To create this plant location database, we downloaded County Business Pattern (CBP) data by sector and county to obtain the number of plants in each county. From the universe of U.S. counties, those without a plant in the desired sector were dropped. Next, we point-located each plant and found the distance to each plant's nearest neighbor. Because CBP data does not provide the longitude or latitude of the establishments, we used the longitude and latitude of the county centroids to estimate the point locations of plants. From each plant site (county centroid), we calculated the crow-fly mile distances to all other plant sites (other county centroids). From this distance matrix we identified nearest neighbor plant pairs and an estimate of the distance between them. The distance between them is the continuous and non-normalized measure of our explanatory variable x_{ie}.

For plants in counties with more than one plant in the sector in the same county, the plants in each pair are classified as "concentrated" ($x_{ie} = 0$). In the case of an odd number of plants, the number of plants in the county was divided by two and then rounded up (e.g. where there are five plants in a county, we included three plant pairs).

To further distinguish co-located from dispersed plant pairs, we needed to estimate each plant's input market radius (extensive margin), d_e* . First, we estimated the input use of each plant as the plant's share of national employment in the sector (CBP data) multiplied by the sector's total input use (input-output USE table).[5]

Next, we accounted for the possibility that where yields are high, inputs can be obtained from smaller market areas (and vice-versa where yields are low), by controlling for local crop yields. We used the average yield per acre in the state that the plant is located, and calculated the yield per square mile in bushels. Then we converted this to the weight (tons) of inputs produced per square mile around each plant. Dividing the weight (in tons) of input used in each plant by the input weight produced per square mile in the plant's location, we have the number of square miles from which the input is sourced over one year (*area*$_e$). We assume a circular market area (area = $\pi \cdot r^2$) for simplicity, so the extensive margin of each plant's input market area is estimated by $d_e^* = \sqrt{area / \pi}$. For each plant pair there are two estimated extensive margins (d_1^*, d_2^*). For plants to be "dispersed" the estimated distance between them (the distance between the centroids of the counties in which they are located) is larger than the sum of their respective extensive margins.

Figure 5 shows the varieties of plant pairs identified from the various location configurations.

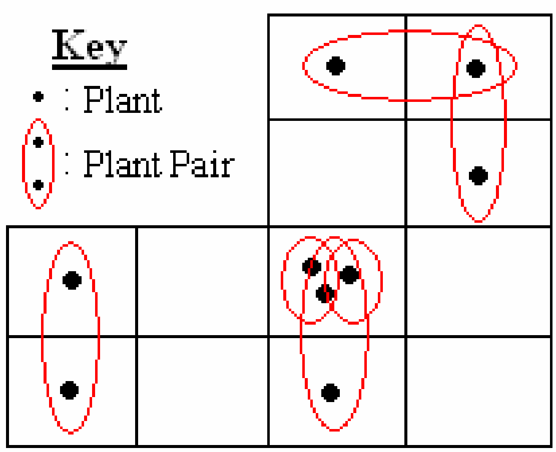


Figure 5. Establishing unique plant pairs

[5] It should be noted that CBP data gives a range of employment for each plant, not exact numbers. A sliding scale was used in the construction of a plant's share to mimic increasing returns to scale: we used the lower limit of the employment range for plants with less than 20 employees and the upper limit for larger plants. That is, we assume that some employees are "fixed cost" types of employees (managers, accountants, janitors…) so that agro-input use increases more than proportionally with plant employment.

The explanatory variables in the regression model of $\boldsymbol{X} = \{x_{ie}\}$ are sector dummy variables $\mathbf{Z} = \{z_i\}$ and the ideal weight ratio * distance compound variables $\boldsymbol{W} = \{w_{ie}\}$. Table 1 presents the preliminary estimates of the model using OLS. The null hypothesis is soundly rejected. The compound ideal weight variable $\boldsymbol{W}$ is significantly negative, as hypothesized, with probability $\alpha \leq .001$.

Table 1. OLS Regression model of the Discrete Dependent Variable x_{ie}

variable		Unstandardized coefficient	Std. error	Standardized coefficient	t ratio	significance
constant		1.857	0.085		21.7	.000
w_{ie}		**-0.0046**	**0.005**	**-0.455**	**-8.87**	**.000**
Zi Sector dummies	Dogfood	-0.323	0.122	-0.128	-2.64	.008
	Flour	-0.451	0.094	-0.226	-4.78	.000
	Malt	-0.217	0.194	-0.038	-1.12	.263
	WetCorn	-0.154	0.152	-0.036	-1.02	.309
	Soy	0.370	0.111	0.109	3.32	.001
	Sugar	-0.398	0.211	-0.072	-1.89	.059

6. CONCLUSIONS

We have shown that agro-industry, an industry whose major input is ubiquitously supplied, and where the price of that input declines with distance from the city; may profitably concentrate, co-locate, or disperse, depending on relative ideal weights. We have clarified why the exclusion theorem does *not* apply for such establishments. When input prices decline with distance from the center, transport cost rates are non-decreasing in distance, and the extensive margins of the input market area are endogenous, we show that costs are convex (profits are concave) with respect to distance from the market center.

Our analysis showed that the initial plant in any agro-industry most profitably locates in the middle of the industry's input supply region, regardless of the relative costs of output and input transport, or 'orientation.' Evidence consistent with this hypothesis has been provided by Kim (1999) and numerous geographic historians. This does not imply, however, that agro-industry is rural, because most places that were/are the most accessible to large supply regions, where agro-industry first appeared, have since become cities (e.g., Beijing, China; Mexico City; Mexico; Toulouse, France; Kansas City, USA). This alone rationalizes the observed preponderance of agro-industrial establishments in urban and metro counties rather than rural ones in the United States.

Our analysis also provided the first testable version of the canonical location-theoretic hypothesis that plant location systematically depends on relative ideal weights. We tested and confirmed this theory directly. Prior empirical analyses of industry location have been *ad hoc* or hedonic models, not direct tests of choice implied by the FOCs of an optimization model. In existing empirical studies, the unit of observation is an area, such as a state or county, and the dependent variable is the number of establishments or the change in the number of establishments in the area. The explanatory variables for those models are characteristics of the areas that are chosen. The implicit assumption tested is that the area is chosen more often if its characteristics are preferred to the characteristics of other sites.

In contrast, we have directly tested establishment location theory and shown it's positive validity. This is important because it justifies the normative math programming models that are widely used to site agro-industrial plants in practice. One of the innovations that enabled us to empirically fit our model's choice rule was our definition of the choice variable. Optimal location problems are usually formalized in terms of exogenously determined market and input site locations, and the choice variable is expressed in terms of the distance of the plant from one of those locations. One difficulty is how can the distance of any one plant from "the" market site or "the" input site be measured? In contrast, we formalized the choice variable in terms of the distance between two plants, which not only measurable but also unique for each plant pair. To reduce the statistical noise arising form the fact that dispersed nearest-neighbor plants in 'the real world' may be hundred or thousands of miles apart, we discretized our dependent variable. The other difficulty is that the constraints T_i, t_i, and q_i do not vary within sectors. There is effectively only one observation per sector. Again, by posing the dependent variable in terms of the distance between nearest-neighbor plant pairs and by taking the local variations in input supply density into account in the construction of d_e*, we were able to obtain as many unique observations per sector as there are plant pairs.

Table 2 presents the model's predictions of the spatial pattern in the locations of agro-industry by sector, according to our estimates of $\hat{x}_i$. If the estimated $\hat{x}_i$ is close to zero (i.e., less than 0.5) the sector is classified as spatially 'concentrated.' If the estimated $\hat{x}_i$ is grater than zero but less than 2, (i.e., $0.5 < \hat{x}_i \leq 1.5$) we classify the sector as 'co-located.' If the estimated $\hat{x}_i$ is about 2 ($\hat{x}_i > 1.5$) we classify the agro-industrial sector as "dispersed."

Table 2. Location Predictions

NAICS	Sector	$\hat{x}_i$	implication	Conner & Scheik classification
311111	Dog and Cat Food Mfg	1.66	Dispersed	Footloose
311211	Flour Milling	1.33	Co-located	Supply-oriented
311213	Malt Mfg	1.57	Dispersed	Supply-oriented
311221	Wet Corn Milling	1.60	Dispersed	Supply-oriented
311222	Soybean Processing	1.66	Dispersed	Supply-oriented
311313	Sugar beet Mfg	1.84	Dispersed	Supply-oriented
311511	Fluid Milk Mfg	1.22	Co-located	Demand-oriented

Our model provides intuitively appealing implications that are generally confirmed by descriptive analysts of agro-industry (e.g. Conner and Schiek); Table 2. It is intuitively reasonable that most of the agro-industrial sectors we modeled are shown to disperse. We suggest equating "dispersed" with Connor & Schiek's classification "supply-oriented." On the basis of ideal weights, we found that only flour milling and fluid milk are expected to locate within each others' input market areas (Table 2). In sum, our model agrees with Connor & Schiek regarding five of the seven industries investigated.

Our model also effectively rationalizes the celebrated case of the flour mill in Atlanta, GA that we assume was sited to profitably process North Dakota wheat. The empirical relevance of our objective classification approach relative to subjective approaches is that a reader of Conner and Schiek would erroneously conclude from their "supply-orientation" classification of Flour mills that the optimal plant location to process North Dakota wheat would be in North Dakota. Using our model one would conclude that a new flour mill would optimally be closer to the destination market, even if there are preexisting flour mills already there. Indeed, there were _n?_ flour mills in Georgia at the time the North Dakota wheat-growing farmer cooperative opened it's new plant in the Atlanta area.

Finally, the constant term in our estimated model is significantly positive. This suggests that the compound effects of missing explanatory variables tends to further disperse plants. It is thus unlikely that we would find a positive effect of local population or other proxies of external agglomeration economies of scale, but that remains to be seen. More fundamentally, an important aspect of our model that remains to elaborated in subsequent research is the endogenization of competition for inputs by neighboring plants on farm output prices. Mill prices for farm output (agro-industrial inputs) should be formally modeled and endogenously determined at each site at the price at which local demand = local supply.

Finally, the micro-spatial model that generates the testable hypotheses applies not only to agro-industry but to any industry that employs spatially dispersed inputs. For example, labor-intensive manufacturing as well as service sectors employ people who must transport themselves to work from their suburban residential areas, and supply products that are distributed from centrally-located transshipment points. Thus this model also holds promise for providing insights about polycentricity, a puzzle that has long preoccupied urban economists.

References

Apland, Jeffrey, and Hans Andersson (1996) "Optimal Location of Processing Plants: Sector Modeling Considerations and an Example," *Review of Agricultural Economics* 18(3):491-504.

Bairoch, Paul (1988) Cities and Economic Development: From the Dawn of History to the Present Chicago: University of Chicago Press.

Barkley, David, and Mark Henry (1998) "Rural Industrial Development: To Cluster or Not to Cluster," *Review of Agricultural Economics* 19(2):308-325.

Beckmann, M. J. & Thisse, J.-F. (1986) The location of production activities. In: Nijkamp P., editor, *Handbook of Regional and Urban Economics*, *Volume 1*. Amsterdam: North-Holland.

Connor, John and William Schiek (1997) Food Processing: An Industrial Powerhouse in Transition, 2nd Edition; New York: John Wiley & Sons.

Dunham, C., R. Sexton, and J.H. Song (1995) "Transportation and Marketing Efficiency in the California Tomato Processing Industry," *Giannini Foundation Research Report 343* University of California, March.

Fuller, S., P. Randolph, & D. Klingman (1976) "Optimizing Subindustry Marketing Organizations: A Network Analysis Approach," *American Journal of Agricultural Economics* 58(3):425-436.

Goetz, Stephen (1997) "State and County-level determinants of Food Manufacturing Establishment Growth: 1987-93," *American Journal of Agricultural Economics* 79(3) August :838-850.

Gruver, G. W. & Xu, D. (1994) "Optimal location and production: a profit function analysis in cartesian space" *Journal of Urban Economics*, **35**:46-70

Henderson, Jason, and Kevin McNamara (1997) "Community Attributes influencing local food processing growth in the U.S. Corn Belt," *Canadian Journal of Agricultural Economics* 45(3):235-250.

Hsu, Song-ken (1997) "The Agro-industry: A Neglected Aspect of the Location Theory of Manufacturing," *Journal of Regional Science* 37(2):259-274.

Hurter, A. & Martinich, J. S. (1989) Facility Location and the Theory of Production. Dordrecht: Kluwer Press.

Kilkenny, M (1998) "Transport Costs and Rural Development," *Journal of Regional Science* 38(2):293-312.

------------ & Thisse, J-F (1999) "Economics of Location: A Selective Survey" *Computers and Operations Research* 26 winter:1369-1394.

------------ and Gerald Schluter (2001) "Value-Added Agriculture Policies across the 50 States" *Rural America* 16(1) May *http://www.ers.usda.gov/publications/ruralamerica/ra161/ra161c.pdf*

Kim, Sukkoo (1999) "Regions, Resources, and Economic Geography: Sources of U.S. Regional Comparative Advantage, 1880-1987." *Regional Science and Urban Economics* 29(1):1-32.

King, Gordon, and Samuel Logan (1964) "Optimum Location, Number, and Size of Processing Plants with Raw Product and Final Product Shipments," *Journal of Farm Economics* 46:94-108.

Lederer, Philip & Arthur Hurter (1986) "Competition of firms: discriminatory pricing and location." *Econometrica*, 54(3):623-640.

Mumford, Lewis (1961) The City in History: Its Origins, Its Transformations, and Its Prospects New York: Harbinger.

Sexton, Richard J. (2000) “Industrialization and Consolidation in the U.S. Food Sector: Implications for Competition and Welfare” *American Journal of Agricultural Economics*, 82(5) November:1087-1104

Sveikauskas, L, J. Gowdy, and M. Funk (1988) “Urban Productivity: City Size or Industry Size?” *Journal of Regional Science* 28():185-202.

Takayama, T. and G. Judge (1964) "Spatial Equilibrium and Quadratic Programming," *Journal of Farm Economics* 46(1) February:67-93.

Venables, A.J. (1996) "Equilibrium locations of vertically linked industries," *International Economic Review*, 37:341-359.

Von Oppen, Matthias, and John T. Scott (1976) "A Spatial Equilibrium Model for Plant Location and Interregional Trade," *American Journal of Agricultural Economics* 58(3):437-45.