An Alternative Theory of the Plant Size Distribution with an Application to Trade

by

Thomas J. Holmes and John J. Stevens[1]

February 2010

[1]Holmes: University of Minnesota, Federal Reserve Bank of Minneapolis, and NBER. Stevens: The Board of Governors of the Federal Reserve System. The views expressed herein are solely those of the authors and do not represent the views of the Federal Reserve Banks of Minneapolis, the Federal Reserve System, or the U.S. Bureau of the Census. The research presented here was funded by NSF grant SES 0551062. We thank Brian Adams, Steve Schmeiser, and Julia Thornton for their research assistance for this project. We thank Shawn Klimek for his help with the Census Micro Data. The statistics reported in this paper derived from Census micro data were screened to ensure that they do not disclose confidential information.

1 Introduction

There is wide variation in the sizes of manufacturing plants, even within the most narrowly-defined industry classifications used by statistical agencies. For example, in the wood furniture industry in the United States (NAICS industry code 337122), one can find plants with over a thousand employees and other plants with as few as one or two employees. The dominant theory of such size differentials, like these that occur *within-industry*, models plants as varying in terms of productivity. See Lucas (1978), Jovanovic (1982), and Hopenhayn (1992). In this theory, some plants are lucky and draw high productivity at startup, others are unlucky and draw low productivity. The size distribution is driven entirely by the productivity distribution.

The approach has been extremely influential. It underpins recent developments in the international trade literature. Melitz (2003) and Bernard, Eaton, Jensen, and Kortum (BEJK) use the approach to explain plant-level trade facts. In Melitz, plants with higher productivity draws have large domestic sales and also have the incentive to pay fixed costs to enter export markets. In this way, the Melitz model explains the fact—documented by Bernard and Jensen (1995)—that large plants within narrowly-defined industries are more likely to be exporters than small plants. Relatedly, in BEJK, more productive plants have wider trade areas. Both the Melitz and the BEJK theories have a sharp implication about how increased exposure to import competition impacts a domestic industry. The smaller plants in the industry—which are the low productivity plants in the industry—are the first to exit.

In our view, the dominant approach goes too far in attributing all differences in plant size within narrowly-defined Census industries to differences in productivity. It is likely that plants that are dramatically different in size are doing different things, even if the Census happens to put them in the same industry. Moreover, these differences in function may be systematic and may very well be directly related to how increased import competition would impact the plants.

Take wood furniture. The large plants in this industry with more than 1,000 employees are concentrated in North Carolina. These plants make the stock kitchen and bedroom furniture pieces one finds at traditional furniture stores. Also included in the Census classification are small facilities making custom pieces to order, such as small shops employing Amish skilled craftsman. Let us apply the standard theory of the size distribution to this industry. Entrepreneurs entering and drawing a high productivity parameter open up megaplants in North Carolina; those with low draws perhaps open Amish shops. The Melitz model and the Eaton Kortum model both predict the large North Carolina plants will have large market areas, while the small plants will tend to ship locally. So far so good, because this is consistent with the data as we show. But what happens when China enters the wood furniture market in a dramatic fashion as has occurred over the past ten years? While all of the U.S. industry will be hurt, the Melitz and Eaton and Kortum theories predict the North Carolina industry will be relatively less impacted because it is home to the large, productive plants. In fact the opposite turned out to be true.

Our theory takes into account that typically in any industry there tends to be some segment providing speciality goods, often custom-made goods, the provision of which is often facilitated by face-to-face contact between buyers and sellers. This speciality segment is the province of small plants. Large plants tends to make standardized products. Here we follow the ideas of Piore and Sable (1984) and a subsequent literature distinguishing between the mass production of standardized products taking place in large plants and the craft production of speciality products taking place in small plants. When China enters the wood furniture market, naturally it enters the standardized segment of the market, following its comparative advantage, making products similar to the stock furniture pieces produced in North Carolina. In this theory, the North Carolina industry is hurt the most, as actually happened.

Our starting point is the Eaton and Kortum (2002) model of geography and trade as further developed in Bernard, Eaton, Jensen, and Kortum (2003) (BEJK). In its basic form, plants vary in productivity and location, but are otherwise symmetric in terms of transportation costs and underlying consumer demand. We take this model "off the shelf" as our model of the standardized segment of an industry, and we fold in a simple model of a speciality segment. We explore two issues in the model. First, how is the size distribution of plants connected to the geographic distribution of plants (call this the plant size/geographic concentration relationship)? Second, if there is a surge in imports, what is the relative impact of the trade shock across locations that vary by geographic concentration and mean plant size?

We estimate the model separately for individual industries, using Census data that includes survey information on the origins and destinations of shipments (the Commodity Flow Survey). The shipment information is critical for our analysis because it enables us to recover parameters related to the transportation cost structure in the BEJK framework.

We obtain three main empirical results. First, we estimate that in most industries, more than half of the plants in an industry can be classified as being in the speciality-goods segment; i.e. this segment dominants plant counts. The second finding relates to the ability of the model to fit the observed plant size/geographic concentration relationship. In the High Point area where the wood furniture industry concentrates, average plant size is four times the national average. A similar pattern holds widely in the data. We find that a constrained version of the model that only allows for the standardized sector can fit the *qualitative* pattern but it fails as a *quantitative:* the predicted differences are too small. The model fits the data once the speciality goods segment is brought in. Variations in average plant size across locations are primarily driven by variations in the share of speciality-good plant counts. The third result concerns the impact of a surge in imports from China. In the constrained model, locations where an industry concentrates that have large plants are predicted to increase their domestic shares after such a trade shook. The opposite happened.

We find it particularly revealing to analyze what has happened in large metropolitan areas. Generally speaking, in recent decades large cities have not been home to huge manufacturing plants in the kinds of industries, like furniture and clothing, that China now dominates. In the United States, large plants in these industries have been concentrated in smaller, manufacturing-oriented areas, like High Point. We show that over the period 1997 to 2007, in industries where exports from China have surged, the domestic industry has shifted towards large metropolitan areas like New York and Los Angeles, places where average plant size has typically been small. These are places where we expect to see a large demand for speciality and custom goods. And we also expect to find there a large supply of inputs suited for speciality and niche products. These are different from the low skill inputs used in mass production of standardized products in large plants—inputs readily available in China and places like Highpoint. Our theory with a specialized good sector can account for how import pressure from China shifts industries from places like Highpoint to places like New York City. Without the specialized-good segment in the model, the shift goes the other way. The early case study of Hall (1961) of the garment industry circa 1960 is interesting to bring up at this point. It explains how the large plants in places like North Carolina were tending to mass produce the standardized garments like nurses uniforms while the small plants in New York city tended to produce fashion items. The new development is that China has entered to play the role of North Carolina, while New York still plays New York (albeit in a relative sense given the overall decline of manufacturing).

There is an emerging new literature that allows for richer forms of heterogeneity across plants than the first generation of trade models with heterogeneous firms found in Melitz (2003) and BEJK. Hallak and Sivadasan (2009) allow plants to differ in the standard way regarding cost structure, but also in a second dimension in terms the plant's ability to provide quality. Their theory can explain why sometimes smaller plants export more than larger plants. (This can happen in their model if the smaller plant has sufficiently higher quality). Bernard, Redding and Schott (2009) develop a multi-product model of a firm with differences not only in an overall firm productivity levels, but heterogeneity in product-specific attributes as well. Our paper is in the spirit of these papers it that it allows for richer heterogeneity. One difference is that we are adding heterogeneity within narrowly-defined industries in the extent to which the goods being produced are tradeable, with customized versions of goods being more difficult to trade. Holmes and Stevens (2004) include a margin like this in regional model linking plant size and geographic concentration. This paper is different from our earlier paper because it (1) uses BEJK to develop a entirely different modeling structure, (2) takes the model to the data and estimates its parameters, and (3) examines the impact of a trade shock.

2 Theory

The first part of this section presents the model. The second part derives analytic results.

2.1 Model

There is a fixed set of L locations, indexed by ℓ. Each location will typically produce goods in a variety of industries. When we go to the data, we will take into account that industries differ in their model parameters. Here we describe the model in terms of a particular industry and leave implicit the industry index.

Consumers have Cobb-Douglas utility function for industry composites. Assume ξ is the spending share on the particular industry we are looking at. For this industry, let p_ℓ be the composite industry price index and q_ℓ be the composite industry quantity. Given the Cobb-Douglas assumption, spending $x_\ell = p_\ell q_\ell$ on the industry at location ℓ equals

$$x_\ell = \xi I_\ell,$$

given income I_ℓ.

The industry has two segments, the *standard* segment indexed by "s" and the *speciality* segment indexed by "b" (where "b" denotes "boutique.") The industry composite q_ℓ is made up of a standard good q_ℓ^s composite and a speciality (or boutique) good composite q_ℓ^b in the usual CES way,

$$q_\ell = \left(\zeta^s \left(q_\ell^s \right)^{\frac{\rho-1}{\rho}} + \xi^b \left(q_\ell^b \right)^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}}, \tag{1}$$

where ρ is the elasticity of substitution between the two segments and the segment weights sum to one, $\xi^s + \zeta^b = 1$. We next describe each segment in turn.

2.1.1 The Standard Good Segment

We use the BEJK model as our model of the standard segment. There is a continuum of differentiated standard goods indexed by $j \in [0, 1]$. For example, if the industry is the wood furniture industry, then j specifies a particular kind of wood furniture, such as a kitchen table of a particular size, a particular finish, a particular shape, and a particular kind of wood. The different standard goods j are aggregated to obtain the standard segment composite q_ℓ^s in the usual CES way. Let σ be the elasticity of substitution and let $P_\ell(j)$ be the price of good j at location ℓ. (For simplicity we leave the "s" superscript implicit here as the j index only refers to tradable goods.) Then the expenditure at location ℓ for good j equals

$$X_\ell(j) = x_\ell^s \left(\frac{P_\ell(j)}{p_\ell^s} \right)^{1-\sigma}$$

where x_ℓ^s is spending on the standard segment composite at location ℓ and p_ℓ^s is the price index,

$$p_\ell^s = \left[\int_0^1 P_\ell(j)^{1-\sigma} dj \right]^{\frac{1}{1-\sigma}}.$$

As in BEJK, there are potential producers at each location with varying levels of technical efficiency. Let $Z_{ki}(j)$ index the efficiency of the kth most efficient producer of good j located at i. This index represents the amount of good j made by this producer, per unit of input.

There is an "iceberg" cost to ship tradable segment goods across locations. Let $d_{\ell i}$ be the amount of good that must be shipped to location ℓ from location i in order to deliver one unit. There is no transportation cost for delivering to the location where the good is produced, i.e., $d_{ii} = 1$. Otherwise, $d_{\ell i} \geq 1$, for $\ell \neq i$. Assume that the triangle inequality

$d_{\ell i} \leq d_{\ell k} d_{ki}$ holds.

The distribution of efficiencies are determined as follows. Let w_ℓ denote the wage at location ℓ, and let T_ℓ denote a parameter governing the distribution of efficiency of the standard segment at location ℓ. Suppose the maximum efficiency Z_{1i} is drawn according to

$$F_i(z) = e^{-T_i z^{-\theta}}.$$

The parameter θ governs the variance of productivity draws.

Eaton and Kortum (2002) show that for a given standard segment good j, the probability location i is the lowest cost producer to location ℓ is

$$\pi_{\ell i} = \frac{\gamma_i a_{\ell i}}{\sum_{k=1}^{L} \gamma_k a_{\ell k}}, \tag{2}$$

for

$$\begin{aligned} \gamma_i &\equiv T_i w_i^{-\theta} \\ a_{\ell i} &= (d_{\ell i})^{-\theta}. \end{aligned} \tag{3}$$

We refer to γ_i as the *cost efficiency index* for location i and $a_{\ell i}$ as the *distance adjustment* between ℓ and i. Let $\Gamma = (\gamma_1, \gamma_2, ... \gamma_L)$ be the cost efficiency vector and A (with elements $a_{\ell i}$) be the distance adjustment matrix. We can think of $\gamma_i a_{\ell i}$ as an index of the competitiveness of origin i at destination ℓ. It starts with location i's overall cost efficiency and adjusts for distance to ℓ. Location i's probability $\pi_{\ell i}$ of getting the sale at ℓ equals its own competitiveness at ℓ relative to the sum of all the other locations' competitiveness at ℓ.

BEJK consider a rich structure with multiple potential producers at each location who each get their own productivity draws. Then firms engage in Bertrand competition for consumers at each location. The equilibrium may feature limit pricing, where the lowest cost producer matches the second lowest cost. Or the lowest cost may be so low relative to rivals' costs that the price is determined by the inverse elasticity rule for the optimal monopoly price. The very useful result of BEJK is that allowing for all of this does not matter. Conditional on a location i landing a sale at ℓ (i.e. that location i is the low cost producer for ℓ), the distribution of prices to ℓ is the same for all originations i. This implies that the sales revenues from ℓ are allocated according to $\pi_{\ell i}$. That is, total sales revenue of origin i to destination ℓ is

$$y^s_{\ell i} = \pi_{\ell i} x^s_\ell.$$

Total sales revenues on standard segment goods originating in i across all destinations is

$$y^s_i = \sum_{\ell=1}^{L} \pi_{\ell i} x^s_\ell.$$

Like BEJK, we associate a plant with a particular good j produced at i. The measure of standard segment goods produced at i equals π_{ii}, the measure of goods location i sells to itself.[2] We allow for a scaling factor ν^s, so that the number of standard segment plants at location i is

$$n^s_i = \nu^s \pi_{ii}. \quad (4)$$

2.1.2 Speciality Segment

We allow for the speciality segment to be modeled in two different ways. Conceptually these two cases are very different. Yet for much of what we do in this paper, the results for the two cases are similar.

Our first case is the *speciality-segment-as-nontradeable-goods* model. As an example of this case, consider the wood furniture industry. This industry includes plants that look like retail stores in a shopping center. A consumer can go into such an establishment and meet face to face with designers to come up with a design for a custom piece. When the furniture is actually made on the premises, the Census classifies the establishment as being a wood furniture manufacturer. This kind of speciality establishment is aimed at a local market, as consumers do not want to drive long distances to meet with a designer. For simplicity, we assume for this case that the transportation cost across locations is infinitely high precluding trade. For this first model then, total sales by specialty segment plants (also called boutique plants) at i equals expenditure at i,

$$y^b_i = x^b_i.$$

Next we simplify by assuming average plant size in terms of sales volume is constant across locations and equals $\bar{s}^b$. Define $\nu^b \equiv 1/\bar{s}^b$ as inverse size. The number of speciality plants

[2] On account of the triangle inequality $d_{\ell i} \leq d_{\ell k} d_{ki}$, if a particular plant is the most efficient producer at any location, it is also the most efficient producer at its own location

at i equals

$$n_i^b = \frac{y_i^b}{\bar{s}^b} = \nu^b x_i^b. \tag{5}$$

The idea here is that there tends to be some efficient size of retail-like, speciality establishments. If speciality-good expenditure doubles at a location, all expansion of the industry at the location occurs on the extensive margin of a doubling of the number of establishments, rather than any increase in size of establishments. Implicitly, there are diseconomies of scale when plants get too big. This is plausible for retail-like, custom operations.

Our second case is the *speciality-segment-as-high-end-niche* model. For this, we go the other extreme and treat the speciality goods as being perfectly tradable. The products have high value to weight and if face-to-face contact between buyer and seller is not important transportation costs are then immaterial. The price p^b the these goods is the same at all locations i. The amount of high-end niche activity at a location depends the supply of factors specific to the segment at the location. We think of this as creative talent or artisanship that is unrelated to the factor T_i determining suitability for standardized goods. The total value of production y_i^b at i depends implicitly on the supply of creativity. Again, assume average plant sales volume is constant at $\bar{s}^b$ across locations with inverse $\nu^b \equiv 1/\bar{s}^b$. The number of speciality plants at origin i then equals

$$n_i^b = \nu^b y_i^b. \tag{6}$$

This high-end-niche model of the speciality segment can be regarded as the limiting case of the BEJK model where transportation cost is zero (or $d_{\ell i}^b = 1$). As we will see below, average plant size is constant across locations in this limiting case of BEJK. In particular, as comparative advantage increases, the resulting expansion of output is met entirely on the extensive margin of more plants.[3]

We limit ourselves to these two extreme cases for technical tractability. We expect that in many cases the speciality good sector will be some combination of these two extreme cases. That is, it will have a hard-to-trade element (because face-to-face contact between the buyer and producer is desirable) and a high-end fashion element (because comparative advantage for the segment depends upon the supply factors like creative talent that is different from

[3]It is worth noting that the nontradeable model of the speciality segment is not a limit case of the BEJK model where $d_{\ell i}^{spec} = 0$. In that limit case, as local demand expands, sales volume expands on the intensive margin of larger average size plants. In our model of the nontradeable segment, an expansion of demand is met on the extensive margin of more plants.

the supply of factors used to produced standardized goods).

2.2 Results

This subsection uses the model to examine two issues. First, how is average plant size at a location related to the concentration of industry at a location? Second, how does an import surge impact the distribution of domestic production? We begin the analysis discussing what happens with only standardized goods. Then we discuss how adding speciality goods impacts the results.

2.2.1 The Plant Size/Geographic Concentration Relationship

To relate average size to industry concentration, we use the location quotient at i to measure industry concentration at i. Recall that y_i is total sales volume of producers located at i and x_i is total expenditure of consumers located at i. Letting y and x be the aggregate totals, the sales location quotient Q_i^{sales} is a location's share of the value of production over its share of expenditure, i.e.,

$$Q_i^{sales} \equiv \frac{y_i/y}{x_i/x}. \tag{7}$$

If the distribution of production exactly follows expenditure, it equals one everywhere and no locations specialize in the industry. Otherwise, if there are locations where this is greater than one, we say the location specializes in the industry and is a net exporter.

Following Holmes and Stevens (2002), we can think of there being two margins over which a location can specialize in an industry: the extensive margin of more plants and the intensive margin of higher average plant size. To highlight these two margins, we decompose the sales location quotient as the product of a *count* quotient and a *size* quotient,

$$\begin{aligned} Q_i^{sales} &\equiv \frac{y_i/y}{x_i/x} = \frac{n_i/n}{x_i/x} \times \frac{y_i/n_i}{y/n} && (8) \\ &= Q_i^{count} \times Q_i^{size}. && (9) \end{aligned}$$

recalling that n_i is the plant count at i and letting n be the aggregate plant count. The count quotient is a locations's share of plant count relative to its expenditure share. The size quotient is a location's average plant size (in sales) relative to the aggregate average plant size.

For now, suppose there is only a standardized good segment so that model reduces to off-the-shelf BEJK. For the benchmark case with no transportation costs we have

Proposition 1. With only standardized goods and no transportation costs ($d_{\ell i} = 1$ all $\ell \neq i$), then $Q_i^{size} = 1$ for all i so average plant size (in sales volume) is identical at all locations. All variation in Q_i^{sales} is through the extensive margin Q_i^{count}.

Proof. With $d_{\ell i} = 1$ all $\ell \neq i$, the probability i serves ℓ in (2) reduces to

$$\pi_{\ell i} = \frac{\gamma_i}{\sum_k \gamma_k}$$

which is independent of destination ℓ. From BEJK this also equals sales share at each location. So average plant size at each location is

$$\bar{s}_i = \frac{\sum_\ell \pi_{\ell i} x_\ell}{n_i} = \frac{\sum_\ell \pi_{\ell i} x_\ell}{\nu^s \pi_{ii}} = \frac{\sum_\ell x_\ell}{\nu^s} \tag{10}$$

which is constant across locations. *Q.E.D.*

Following intuition about the Eaton Kortum setup in Alverez and Lucas, the productivity at a location has an interpretation of being a function of the first-order statistic of draws from the exponential distribution. A location i with productivity T_i twice as high as another location can be interpreted as having twice as many underlying draws. With transportation costs that are zero, it is intuitive that a location with twice as many underlying draws will produce twice as many products.

When transportation cost is positive, the BEJK model can deliver differences in average plant size across locations. Analytical results are difficult to come by, but some basic patterns can be readily discerned by numerical examples. Table 1 illustrates a numerical example with two locations. For all the parameters considered, $Q_2^{sales} = 4$, so location 2 specializes in the industry to a substantial degree. For location 1, $Q_2^{sales} < 1$, and the extent to which this is less than one depends upon assumptions related to the distribution of expenditure x_1 and x_2 across locations. We consider three possibilities for the expenditure distribution: in the first, expenditure is equal across the two locations, in the second it is three times higher in location 2, in the third it is three time higher in location 1. We also vary the distance adjustment parameter $a_{12} = (d_{12})^{-\theta}$. For different levels of a_{12}, we back what the ratio γ_2/γ_1 must be that would result in $Q_2^{sales} = 4$ and then calculate the corresponding Q_2^{count} and Q_2^{size}.

We start by discussing the equal expenditure case. Note that as a_{12} is decreased as we move down the table (so the distance adjustment becomes more important), the productivity advantage of location 2 (γ_2/γ_1) must be increased to hold constant the net trade between the two locations (i.e. to keep $Q_2^{sales} = 4$). At the limiting case where $a = 1$, average size is the same in both places, $Q_2^{size} = 1$, consistent with Proposition 1. For this limiting case, the expansion in sales at location 2 comes entirely through the count margin, $Q_2^{sales} = Q_2^{count}$. As a_{12} is decreased, average size in location 2 increases, so both the size margin and the count margin play a role. But even as a shrinks to extreme levels (and the implied γ_2/γ_1 goes to extreme levels), the establishment margin is always greater than the size margin. This holds more generally in other numerical examples with $x_1 = x_2$. This discussion gives a sense of how the the BEJK model has trouble accounting for large differences in average plant size across locations, particularly if transportation costs are relatively small.

If expenditure is larger in location 2 and if trade is sufficiently difficult, the size margin will be significant. It is intuitive that when trade is difficult in the BEJK model, plants in locations with high local expenditure will tend to be big, as most sales are local. In what we discuss below, high local demand will not be a relevant explanation for the large average size plants to be found in industrial centers like Highpoint, North Carolina. There places are relatively small cities with low local demand.

2.2.2 Response to an External Trade Shock

We next examine how a trade surge in the standardized segment impacts the distribution of the production of standardized goods across locations. We model trade in a simple fashion. Suppose as above there are two domestic locations. Now add a third location that we call location C (China). We assume location C does not have any expenditure, $x_C = 0$. Assume the transportation cost is the same from location C to both domestic locations, $d_{1C} = d_{2C} = d_C > 1$ and let $a_C = d_C^{-\theta}$. Let $\lambda \equiv a_C \gamma_C$ be C's competitiveness index (which is the same at locations 1 and 2). Then extending (2), the probability location 2 sells at 1 is

$$\pi_{12} = \frac{a_{12}\gamma_2}{a_{11}\gamma_1 + \alpha_{12}\gamma_2 + \lambda} \tag{11}$$

and the general formula for $\pi_{i\ell}$ is analogous. In the location quotients below, only domestic sales are used to calculate sales share (the aggregate y excludes sales from C).

Proposition 2. Suppose there are only standardized goods. Suppose the parameters are such that location 2 is more cost efficient than location 1, $\gamma_2 > \gamma_1$, that expenditures are the same $x_1 = x_2$, and that there is some distance adjustment $a_{12} < 1$. (i) Then location 2 is the *high concentration location*,

$$\begin{aligned} Q_2^{sales} &> 1 > Q_1^{sales}, \\ Q_2^{count} &> 1 > Q_1^{count}, \\ Q_2^{size} &> 1 > Q_1^{size}. \end{aligned}$$

(ii) The sales location quotient Q_2^{sales} strictly increases in the competitiveness index of location C.

Proof. See appendix.

The expansion of the foreign location C hurts sales at both locations. But it hurts sales relatively less in location 2 where the industry is concentrated.

We note that our result is a partial equilibrium result for a particular industry. When we increase competitiveness λ of C in this industry, we are holding fixed wages w_i at each location.

2.2.3 How Introducing the Speciality Sector Changes the Results

It is straightforward to see how introducing the speciality sector changes the results. Consider first average plant size at a location. It equals the weighted average of the mean plant size in the standard segment and speciality segments.

$$\begin{aligned} \bar{s}_i &= \frac{y_i}{n_i} = \frac{n_i^s}{n_i^s + n_i^b}\frac{y_i^s}{n_i^s} + \frac{n_i^b}{n_i^s + n_i^b}\frac{y_i^b}{n_i^b} \\ &= \frac{n_i^s}{n_i^s + n_i^b}\bar{s}_i^s + \frac{n_i^b}{n_i^s + n_i^b}\bar{s}^b \end{aligned}$$

It is plausible that average size of standardized plants $\bar{s}_i^s$ is typically much larger than the average size $\bar{s}^b$ of speciality plants. So differences in mean plant size across locations can be driven by differences in the composition of types of plants. Highpoint can have a large average plant size if it has a large share of standardized plants.

Next consider the impact of the trade shock. For simplicity, assume the elasticity of substitution ρ between the two segments equals one (Cobb-Douglas) fixing the spending shares on the two segments. To discuss the impact of the shock, we need to distinguish between the two models of the speciality sector that we have put forth. We begin with the nontradable case. The emergence of the new foreign location C is irrelevant for the nontradable speciality sector because the infinitely high transportation costs keeps any goods from C outside of this segment. Suppose there are two domestic locations as in Proposition 2 and the standardized segment lies completely within location 2. The nontradable segment is distributed across the two locations following expenditure. Assume the Census combines the standardized and speciality segments into one combined industry. Then location 2 will be measured as specializing in the overall industry. If average plant size of standardized plants is greater than speciality plants, then location 2 will have larger plants than the domestic average. So part (i) of Proposition 2 continues to hold. But now consider the trade shock from the increase in comparative advantage of location C. This displaces sales of standardized goods at location 2, but has no impact on sales of speciality goods at location 1 (the only type of products sold at 1). Thus part (ii) of Proposition 2 does not hold here. Location 2 with high industry concentration and the large plants loses share relative to location 1 on account of the surge from location C.

In the alternative model where the speciality goods are high-end niche goods, but very tradable, the outcome depends upon the emerging trade partner's ability to compete in the speciality segment as well as the standardized segment. It is likely that the circumstances that make location C a strong competitor in the standardized segment (e.g. an abundance of unskilled labor) are unrelated to its competitiveness in the speciality segment. If that is the case, the speciality segment is not impacted and part (ii) of Proposition 2 won't hold in this case either.

The discussion so far imposes Cobb-Douglas utility so spending on speciality goods stays fixed. Suppose instead that the elasticity of substitution ρ between speciality and standardized goods is greater than one. But also assume that the elasticity σ cross standardized goods exceeds ρ. So standardized goods are better substitutes for each other than speciality goods. Then the advent of location C in standardized goods will carry over to the speciality goods, as consumers substitute high price niche goods with low price standardized goods. Nevertheless, the standardized sector is impacted relatively more and Proposition 2 part (ii) goes the other way. The location with high market share and large plants loses domestic market share.

3 The Data and some Descriptive Results

The first part of this section discusses data sources and industry and geographic classifications. The second part provides some initial descriptive results.

3.1 The Data

We analyze the confidential micro data for two programs of the U.S. Census Bureau. The first is the 1997 *Census of Manufactures* (CM), The data are collected at the plant level, e.g. at a particular plant location, as opposed to being aggregated up to the firm level. For each plant, the file contains information about employment, sales, location, and industry classification.

The second data file is the 1997 *Commodity Flow Survey* (CFS). The CFS is a survey of the shipments that leave manufacturing plants. See Hillberry and Hummels (2008) for details about these data. Respondents are required to take a sample of their shipments (e.g. every 10 shipments) and specify the destination, the product classification, the weight, and the value of the shipment at origin. On the basis of this probability weighted survey, the Census tabulates estimates of figures such as the total ton miles shipped of particular products. There are approximately 30,000 manufacturing plants with shipments in the survey. Holmes and Stevens () provides more details..

While we have access to the raw confidential Census data, in some instances, we report estimates based partially on publicly-disclosed information rather than entirely on the confidential data. These are cases where we want to report information about narrowly-defined geographic areas, but strict procedures relating to the disclosure process for the micro-data based results get in our way. In these cases, we make partial use of the detailed public information that is made available about each plant in the Census of Manufactures. Specifically, the Census publishes the cell counts in such a way that for each plant, we can identify its six-digit NAICS industry, its location, and its detailed employment size class (e.g., 1-4 employees, 5-9 employees, 10-19 employees, etc.). We use this and other information to derive sales and employment estimates for narrowly-defined geographic areas. The data appendix provides details.

Plants are classified into industries according to the *North American Industry Classification System* or NAICS. The finest level of plant classification in this system is the six-digit level and there are 473 different manufacturing industries at this level. When we estimate the model, we focus on a more narrow set of 172 manufacturing industries. These are industries with diffuse demand that approximately follows the distribution of population. We do this so we can use population to proxy demand when we estimate the model.[4] Specifically, through use of the input-output tables, we selected industries that are final goods for consumers. In addition, we included intermediate products used in things like construction and health services that have diffuse demand. We excluded intermediate products used downstream for further manufacturing processing. See the data appendix for additional details.

We use Economic Areas (EAs) as defined by the Bureau of Economic Analysis (BEA) as our underlying geographic unit. There are 177 EAs that form a partition of the contiguous United States. The BEA defines EAs to construct meaningful economic geographic units, using counties as building blocks. A metropolitan statistical area (MSA) is typically an EA. In addition, rural areas not part of MSAs get grouped into an EA. To calculate distances between locations, we take the population centroids of each EA and use the great circle formula.

4 Some Descriptive Results

This subsection presents descriptive evidence that sheds light on the plausibility of our thesis. We begin by looking at a selected set of seven industries for which we are able to exploit additional information about what the plants do beyond the NAICS code. We then discuss some preliminary findings for a broader set of industries.

In 1997, the Census changed its industry classification from the SIC system to the NAICS system. Seven NAICS manufacturing industries were redefined to include plants that had previously been classified as retail under SIC. For example, under the SIC system, establishments that manufactured chocolate on the premises for direct sale to consumers were classified as retail. Think here of a fancy chocolate shop making premium chocolate by hand. These were moved into NAICS 311330, "Confectionery Manufacturing from Purchased Chocolate." This industry also includes chocolate factories with more than a thousand employees. This situation—where retail candy operations are lumped into the same

[4] There are a variety of ways we can potentially improve our demand measure. For example, for consumer goods industries, we could take into income. For some of our industries that sell investment goods to the general manufacturing sector, we could use the distribution of manufacturing. While our population-based measure is crude, it gets the order of magnitude of demand correct. Iis useful as a first cut in making it possible to consider a wide swath of industries.

industry as mass-production factories making standardized goods—epitomizes what we are trying to capture in our model. Analogous to chocolate, facilities making custom furniture and custom curtains in storefront settings were moved from retail under SIC to manufacturing under NAICS. The logic underlying these reclassifications was an attempt under the NAICS system to use a "production-oriented economic concept" (Office of Management and Budget, 1994) as the basis of industry classification. The concept is that plants that use the same production technology should be grouped together in the same industry. The previous SIC system sometimes followed this logic but was inconsistent in its application.

Table 2 shows the seven NAICS manufacturing industries that were impacted this way. We will refer to these industries as the *1997 Reclassification Industries* and sometimes the *7 industry sample.* All are consumer goods industries (some kind of candy, textiles, or furniture). We do not regard these reclassifications in 1997 as a "mistake" by the statistical authority or any kind of deviation from normal philosophy about how to aggregate plants into industries. It is not feasible for the Census to define industry boundaries too narrowly because otherwise cells become so thin in tabulations that disclosure issues preclude publication. The Census must aggregate in some way and its general procedure is to group standardized versions and speciality versions of the same product into the same industry.

That the Census cleaned up what it considered an earlier mistake is fortunate for our purposes as it yields additional information that can be exploited. The micro data for 1997 contains a plant's SIC code in addition to its NAICS code (because tabulations were published both ways for this switchover year.) We take the sample of seven NAICS industries and refer to the plants that are in retail under SIC as SIC/Retail plants and the remaining plants as SIC/Man. The first thing to note in Table 2 is the large differences in mean plant size between these two groups. The SIC/Retail plants are significantly smaller than the SIC/Man plants.

A well known result due to Bernerd and Jensen is that large plants are relatively likely to export compared to small plants. Table 2 reports exports by type. There is a consistent pattern that the SIC/Man plants (which are large on average) have a 3 percent export share while the SIC/Retail plants (which are small) don't export at all. Furthermore, we can think of retail status as an extreme form of non-exporter status; retailers (typically) do not sell to domestic destinations outside their own immediate vicinity, let alone foreign destinations. So the connection reported here between plant size, export status and retail status can be interpreted as a variant of a well-known empirical pattern.

Suppose we take the speciality-segment-as-nontradeable-good variant of our model. Suppose we take SIC/Retail status as a proxy indicator for our speciality, nontradable good. An immediate implication of the model is that the geographic distribution of the SIC/Retail segment will closely track the distribution of demand. The last column of Table 2 provides evidence consistent with this pattern. It reports a measure of the distribution of industry sales which shows that plants in the SIC/Retail segment tend to closely follow population (as retail does more generally) while plants in the SIC/Man segment tend to be geographically concentrated.

To explain the measure, recall the definition of the location quotient Q_i^{sales} (7) at location i for a given industry, but now use population share to approximate expenditure share in the denominator. Analogous to Holmes and Stevens (2004b), for each industry we sort locations (economic areas) by the location quotient and then aggregate locations into ten approximately equal-sized population-decile classes. (This aggregation helps smooth the data. It is approximate because of lumpiness in location size.) Let Q_d^{sales} be the location quotient of decile d. By definition, $Q_d^{sales} \leq Q_{d+1}^{sales}$. If all sales are concentrated in the top decile then $Q_{10}^{sales} = 10$, as 100 percent of the industry is concentrated among 10 percent of the population.

We are interested in comparing the geographic dispersion of different groups of firms within the same industry. Let g index a particular group of firms. (For example, for what we do in Table 2, the index g signifies whether a plant is SIC/Retail or SIC/Man.) Suppose plants located in decile d of type g are indexed by k and let $y_{d,g,k}$ be the sales of plant k of type g at decile d. Let $\bar{Q}^{sales}$ be the sales-weighted overall mean location quotient across plants from all locations of all types. Then

$$\bar{Q}^{sales} \equiv \frac{\sum_d \sum_g \sum_k y_{d,g,k} Q_d^{sales}}{\sum_d \sum_g \sum_k y_{d,g,k}} = \frac{\sum_d y_d \frac{\frac{y_d}{y}}{\frac{1}{10}}}{y} = 10 \left[\sum_{d=1}^{10} \left(\frac{y_d}{y} \right)^2 \right].$$

So the mean location quotient $\bar{Q}^{sales}$ is exactly the standard Herfindahl index of industry concentration, times a factor of 10. If the entire industry is concentrated in the top decile, then $\bar{Q}^{sales} = 10$. If it is spread equally across the ten deciles then $\bar{Q}^{sales} = 1$. The main interest for this subsection is the *conditional mean location quotient* of plants of type g,

$$\bar{Q}_g^{sales} = \frac{\sum_d \sum_k y_{d,g,k} Q_d^{sales}}{\sum_d \sum_k y_{d,g,k}}.$$

Note that conditioning on type g enters only through the weights; plants of all types in decile d are used to define the Q_d^{sales} associated with a sale.[5] Conceptually, we are taking each dollar of sales in the data and associating it with the location quotient of its origin and taking means.

The last column of Table 2 presents the mean location quotients conditional on SIC/Retail or SIC/Man status.[6] There is a clear pattern in the table that the SIC/Man segments tend to have significantly higher concentration than the corresponding SIC/Retail segments. Moreover, the measures for the SIC/Retail segments are close to one. For example, in the wood furniture industry, the mean is 1.52 for SIC/Retail and 7.11 for SIC/Man.

The SIC/Retail plants in these industries are clearly what we have in mind by speciality plants. But what about the many small plants in the SIC/Man segment in these industries? We can address this by exploiting data on product shipments. The Census of Manufactures asks a sample of plants to itemize their shipments in various product categories. In most cases, the product definitions are unrelated to the *speciality product* versus *standardized product* distinction that would be useful for our paper. However, for the reclassification industries, new product definitions were created as part of the 1997 reclassification that get directly to the core of what we wish to look at. New product categories were created for custom-made furniture products and for candy manufactured for retail sale on premises. We define such products as *speciality* products (see the appendix for the full definition) and determine for each plant with the requisite data the *speciality product share* of shipments. (Not all small plants are required to fill out the detailed survey of shipments and we throw out imputed values, so our data here is for a sample of plants rather than for the universe.) We have usable data for five of the reclassification industries. Table 3 reports unweighted means of this variable for the five industries together as well as three different industry subgroups.

For all five industries together, the mean speciality share is .82 in the SIC/Retail segment and falls to .40 in the SIC/Man segment. Within the SIC/Man segment, mean share falls sharply with plant size, ranging from .57 in the smallest plant category to .07 in the largest. Looking within the more narrow industry groupings, we see the speciality share is uniformly high in the SIC/Retail segment for all narrow industry groupings. Looking within the

[5] In Holmes and Stevens (2002) we calculate analogous measures that for each plant excludes the plant's own contribution to the location quotient and only uses the neighboring plants. This correction makes little difference in what we do here.

[6] Note: in the current version, the geographic partition being used for Table 2 is Census Divisions. Table 4 below uses the deciles described above.

SIC/Man segment, the speciality share is quite high in the wood kitchen cabinet industry. The share falls sustantially with plant size within this segment, .70 to .28. For the SIC/Man plants within the furniture industry, the speciality share also falls sharply with plant size, from .08 to .03.

Naturally, our interest extends beyond the seven reclassification industries in Table 2. Outside these seven industries, we do not have useful product and SIC code distinctions to work with. We do have plant size for each establishment and our last descriptive exercise makes use of it. We break plants down into four employment size categories and calculate the conditional mean location quotient $\bar{Q}_g^{sales}$ for each size category g. This is the column labeled "raw" in Table 3. Next we add six-digit NAICS fixed effects and report how the fitted values vary with plant size, holding industry effects fixed (at the mean level).[7] These go in the column labeled "fixed effects."

As a reference point, we begin with the seven reclassification industries and the results are in the top panel of Table 4. In the column labeled raw, we see the mean location quotient is only 1.36 in the smallest size category and it rises all the way to 6.13 for the largest category. We expect that part of this relationship stems from the fact that some industries (like wood cabinets) tend to have small plants and be geographically disperse. While the inclusion of 6-digit NAICS fixed effects does attenuate the relationship, it remains quite large, going from 2.46 in the smallest plant size category to 5.83 at the top. This pattern is consistent with the pattern established in Table 2 for these industries. There, the large plant category is more geographically concentrated than the small plant category.

In the second panel we do the same exercise for the 165 other Industries with diffuse demand for which we will estimate the model. The same pattern holds. Using fixed effects, mean location quotient increases from 3.72 in the smallest plant size category to 5.41 in the largest category. The spread found here, 3.72 to 5.41 (a difference of 1.69), is half the spread of 2.46 to 5.83 (3.37). This is not surprising as the seven are a selected sample that exemplify that factors we are highlighting. However, the same qualitative relationship that holds in our narrow reclassification sample also holds in the broad sample. We obtain similar results in the bottom panel for the remaining 302 industries that do not have diffuse demand.

[7] We regress plant LQ on the size categories and industry fixed effects, weighting by sales. We then contstruct fitted values by plant size category evaluated at the mean fixed effects.

5 Estimation of the Model

This section estimates the model. The first subsection considers a constrained version of the model where the standardized goods segment is the entire industry. This subsection serves the role of providing first-stage estimates that are used later. The next subsection brings the speciality-goods segment into the analysis.

5.1 First-Stage Estimates: The Constrained Model with Only Standardized Goods

In what we call the first-stage estimates, the model is estimated under the assumption that each six-digit NAICS is a distinct standardized-products industry, i.e. each industry has its own BEJK model parameters. This procedure pins down distance adjustment parameters that will be used throughout the paper.

For each industry h, the data generating process for the industry is summarized by a vector $\Gamma^h = (\gamma_1^h, \gamma_2^h, ...\gamma_L^h)$ that parameterizes the relative productive efficiencies of the various locations and a $L \times L$ matrix A^h, with elements $a_{\ell i}^h$ that parameterize the distance adjustments in (3). We normalize so the γ_i^h sum to one across locations i.

Assume the distance adjustment for industry h takes the form

$$a_{\ell i}^h = a^h(dist_{\ell i}) = \exp(-\eta_1^h dist_{\ell i} - \eta_2^h dist_{\ell i}^2) \tag{12}$$

so that $\ln a_{\ell i}^h$ is quadratic. If $\eta_2^h = 0$, then η_1^h is the decay rate per mile for industry h..

We have restricted attention to the 172 industries listed in the appendix for which demand is diffuse and roughly follows population. For simplicity, we assume this is exactly true so that expenditure x_i^h at location i in industry h is proportional to population and normalize so $\sum_i x_i^h = 1$. The Census of Manufactures (CM) covers the universe of all plants in the United States. Subtracting out the exports of each plant, we can aggregate the plant-level sales data to get y_i^h, the value of domestic shipments originating at location i in industry h.

Other then export information, there is no destination information in the CM. However, the CFS provides survey information on shipments and their destinations that we can link to plants (and thereby determine industry). A concern we have with the CFS data is that local shipments may be over-represented in the data. These seem too high, more than can be absorbed by local demand. We expect that sometimes shipments intended for far away destinations get there by way of a local warehouse In cases like these, the destination found in the CFS may be the local warehouse rather than the ultimate destination. In the appendix, we provide some preliminary evidence of a link between excess local shipments for an industry and an industry's use of the wholesaling sector.

The form of our data leads us to the following strategy for estimating $\eta^h = \left(\eta_1^h, \eta_2^h\right)$ and the productivity vector Γ^h for each industry. We pick (η^h, Γ^h) that perfectly match the distribution of sales at originating locations as we directly observe the universe of sales at each location. Because of our concern about excessive local shipments, we throw out all local shipments in the CFS that are less than one hundred miles and fit the conditional distribution of the longer shipments. Formally, set $\overline{dist} = 100$ and let $B(i, \overline{dist})$ be the set of all destinations at least $\overline{dist}$ from an originating location i. The conditional probability that an industry h shipment originating in location i goes to a particular destination $\ell \in B(i, \overline{dist})$ equals

$$p_{\ell i}^{h,cond} = \frac{y_{\ell i}^h}{\sum_{\ell' \in B(i,\overline{dist})} y_{\ell' i}^h}.$$

For each value of η^h, we solve for the vector Γ^h such that the predicted total sales of the industry at a given location equals total sales in the CM data. We show in the appendix there is a unique solution $\Gamma^h(\eta^h)$ to the 177 nonlinear equations for the 177 locations and derive an iterative algorithm for calculating it. We can then write the conditional probability above as a function of η^h. We pick η^h to maximize the conditional likelihood of the destinations observed in the shipment sample.[8] We provide details about the estimation algorithm in the appendix.

Table 5 reports estimates for several selected industries. The parameter estimates $\hat{\eta}_1^h$ and $\hat{\eta}_2^h$ are in the table, the corresponding 177-element productivity vector $\Gamma^h(\eta^h)$ for each industry can be obtained as a separate data file. The reported industries are those at the 25th percentile points in the distribution of the implied value of $a(100)$, the distance adjustment at 100 miles. The bottom industry in this dimension is "ready mixed concrete," a well-known example of an manufacturing industry in which shipments are overwhelming local. (see Syverson ()). For this industry and four other industries where shipments are overwhelming local (such as "ice" and "concrete blocks") we used all shipments, including those below one hundred miles.[9] We see in the table that the estimate of $a(100)$ for ready mix concrete is

[8] The sample of plants selected for the CFS is stratified. We use the establishment sampling weights to reweight the cell count realizations and follow a pseudo-maximum likelihood approach. In writing down the likelihood, we condition on the origination of a given shipment.

[9] If after exlcuding shipments below 100 miles, the implied value of turned out to satisfy $a(100) \leq .2$,

.01. For ice (not shown in the the table), $a(100) = .06$ and for asphalt paving it equals $a(100) = .09$. These industries are virtually nontradable beyond one hundred miles. Butter is the 25th percentile industry. For this industry there is a high degree of tradability at one hundred miles ($a(100) = .74$), but things drop of steeply at five hundred miles ($a(500) = .27$). The highest ranked tradability industry is "Other Hosiery." We truncate the $a(dist)$ function at one in a few industries like this where the unconstrained value exceeds one. Imposing this constraint makes little difference; unconstrained, the distance adjustment for "other hosiery" at one hundred miles is $a(100) = 1.06$.

Table 5 also reports the mean values of the parameter estimates across all industries. On average, $\eta_1 = .003$ meaning that if we look only at the linear component of (12), the average dropoff in a is .3 percent per mile. The fact that the coefficient η_2 on the quadratic term is negative adds a convexity element to the relationship; the dropoff decreases with distance at a decreasing rate.

We have also reestimated the model using the data from the previous census. We call this the 1992 SIC sample because industry classification was based on SIC that year. We use the same selection criterion to identify industries with diffuse demand and arrive at 175 industries. The mean values of the estimates are virtually the same as in the baseline 1997 NAICS case. For those industries with no change in definition between the 1992 SIC and 1997 NAICS there is a very high correlation in the implied values of $a(dist)$. The CFS survey for the earlier period was a larger, better-funded survey, with many more observations. In particular, the average number of shipments used to estimate the parameters of each industry is 8,500 for the earlier period and about 4,000 in the later period. Thus, estimates for the 1992 SIC sample are more precise.

Table 5 also shows what happens to the estimates when the distance threshold for including shipments in the analysis is varied. As discussed above, CFS manufacturing shipments may overstate local shipments, as some local shipments may end up in the wholesale sector to be ultimately shipped to distance destinations. When we set $\overline{dist} = 0$, so that no observations are excluded, the average value of $a(100)$ falls from the baseline level of .80 to .71. With local shipments included, the model is accounting for the high relative likelihood of local sales by making transportation costs higher. If we go in the other direction and raise the cutoff to $\overline{dist} = 200$, the average value of $a(100)$ rises from .80 to .85. We later discuss the robustness of our main results to choice of $\overline{dist}$.

we reestimated the model with all the shipments and used this estimate instead. For the five industries impacted this way, we constrained $\eta_2 = 0$ and just allowed for linear term η_1.

The last topic of this subsection is goodness of fit. Recall that by construction, the total shipments originating in each location in the estimated model perfectly fits the data. For a notion of goodness of fit, we look at the distance pattern of shipments. We break the shipments above 100 miles into three distance categories, (1) 100 to 500 miles, (2) 500 to 1000 miles and (3) over 1000 miles. For industry h, let s_c^h be the share of the shipments above 100 miles that are in distance category $c \in \{1, 2, 3\}$ in the data. Let $\hat{s}_c^h$ be the fitted value in the estimated model. Table 6 presents descriptive statistics. In the data, on average across the industries, a share .44 of the 100-mile-plus shipments are in the 100 to 500 mile category. This compares to an average share of .38 in the estimated model. The model has a tendency to somewhat understate the shortest distance category and somewhat overstate the two longer distance categories. By construction, the destination of shipments in the model exactly follows the distribution of population. So locations far away from any producers will nevertheless be required in the model to receive their share of shipments. The last part of Table 6 shows that the fitted values of the model do a very good job in accounting for the cross-industry variation in the distance distribution. The slope of s_c^h in a regression on $\hat{s}_c^h$ is approximately one for all three categories.

5.2 Including Speciality Goods

We now introduce the speciality-goods segment into the estimation. It is reasonable to expect that this segment may typically account for a relatively small share of industry sales receipts but while at the same time potentially account for a nonnegligible fraction of establishment counts. This expectation motivates the following estimation strategy. We proceed under the assumption that the standard-segment weight $\xi^{s,h}$ in utility (1) for each industry h is close to one and the specialty-segment weight $\xi^{b,h}$ is close to zero. In this case, the sales distribution by origin and destination is approximately the same as it is in the limiting case where $\xi^{s,h} = 1$ and $\xi^{b,h} = 0$. We then take our estimate (η^h, Γ^h) for each h from above as an approximation to what we need here.

With an estimate of η^h and Γ^h in hand, we can determine the plant counts for standardized goods (subject to the scaling normalization $v^{s,h}$). Recall from (4) this equals $n_i^{s,h} = \nu^{s,h}\pi_{ii}^h(\eta^h, \Gamma^h)$, where we now add a superscript for the industry h. Next consider plant counts for speciality goods. In the nontradable case, using equation (5), speciality plant counts equal $n_i^{b,h} = \nu^{b,h}x_i$. In the high-end niche case, counts will depend on supply of speciality-specific factors through (6). As a first cut, assume supply is proportional to population. This delivers $n_i^{b,h} = \nu^{b,h} x_i$ for this case as well. So for either case, the total number of plants in industry h—standardized plus specialty—equals

$$n_i^h = \nu^{s,h}\pi_{ii}^h(\eta^h, \Gamma^h) + \nu^{b,h} x_i. \tag{13}$$

To take this to the data, we introduce an error term. Suppose the observed total number of plants in industry h at location i equals the above plus an error term $\lambda^h + \varepsilon_i^h$,

$$\tilde{n}_i^h = \nu^{s,h}\pi_{ii}^h(\eta^h, \Gamma^h) + \nu^{b,h} x_i + \lambda^h + \varepsilon_i^h, \tag{14}$$

where ε_i^h is mean zero and has variance proportional to location i's population (i.e. proportional to x_i). We use weighted least squares to estimate for each industry the slopes $\nu^{s,h}$ and $\nu^{b,h}$ and the constant λ^h. (Given the results of the first stage, $\pi_{ii}^h(\eta^h, \Gamma^h)$ is data at this point.)

Table 7 presents the results. The individual estimates are reported for the seven reclassification industries from Table 2; just means are reported for the broader set of industries. We first note that allowing for the constant term λ^h makes little difference; when we reestimate (14) without an intercept we get similar results. Next note that the coefficient estimate $\hat{\nu}^{b,h}$ for speciality goods tends to be quite large. Given the scaling that the x_i sum to one, $\hat{\nu}^{b,h}$ is an estimate of the total count of speciality goods plant in industry h across all locations Define the implied speciality count share to be

$$\text{Speciality Count Share for industry } h \ = \frac{\nu^{b,h}}{\nu^{b,h} + \sum_i \nu^{s,h}\pi_{ii}^h(\eta^h, \Gamma^h)}$$

This model statistic is reported in Table 7 and it is quite high for all the reclassification industries, averaging .75 across the seven industries.

The last two columns of the table report the share of plant counts and the share of sales, for plants with less than 20 employees. We call these *small plants*. We do not necessarily associate all small plants with the speciality goods segment. In the BEJK model, plants can be small if they draw low productivity but have some other advantage that enables them to be open (like the nearest efficient rival being a great distance away). So in our theory, plants can be small because they are speciality plants or because they are low productivity standardized good plants. Even though there is no one-to-one mapping between small plants in the data and speciality plants in the model, we expect there to be some connection so it is useful to relate our estimates to the empirical shares of small plants in the data. In the data for the seven reclassification industries, on average small plants make up a .81 share of establishments but only a .13 share of sales. Small sales shares like these motivate our approach of ignoring the speciality segment when we estimate η^h and Γ^h in the first stage.

Next consider the results for the remaining 165 diffuse demand industries. The R^2 of regression (14) averages .71 across the industries. This is high, even though it is lower then the average of .84 for the reclassification industries. The average speciality plant share of .65 is high, through lower than the .75 average for the reclassification industries. These 165 industries have relatively fewer small plants then the 7 reclassification industries (.57 instead of .81), so is perhaps not surprising that the average estimated speciality plant count comes in lower.

Figures 1 and 2 are a scatterplots across industries with the speciality plant count share in the model on the horizontal axis and the small plant count share in the data on the vertical axis. Figure 1 does this for the 7 reclassification industries. The positive connection is readily evident. The industry with the highest speciality share has the highest small plant share in the data. (In Table 7 we can see this is the curtain industry.) The industry with the lowest speciality share (upholstered wood furniture) has the lowest small plant share. Figure 2 is a plot for the remaining 165 industries. The first thing to note is that the vast majority of industries are on the right side of the graph meaning they have an estimated speciality plant share above .5. Next look at the curve that is fitted by locally-weighted regression. There is an upward slopping relationship with a slope on the order of one half. The correlation coefficient equals .37 (p = .0001). This compares to a correlation coefficient of .77 (p = .05) for the 7-industry sample. Analogous to earlier results, the same qualitative pattern found to hold in the 7-industry selected sample also holds for the broader 165-industry sample, albeit in an attenuated fashion.

It is instructive to examine outliers in the relationship of Figure 2. There are three industries with small plant shares that exceed .70 in which the estimated speciality share is less than half this level. These are:

- *Industrial mold mfg* (In model, speciality count share is.16; in data, small plant count and sales shares are .72 and .23)
- *Special die & tool, die set, jig, & fixture* (In model, speciality count share is.04; in data, small plant count and sales shares are .77 and .25)

- *Costume jewelry & novelty mfg.* (In model, speciality count share is.30; in data, small plant count and sales shares are .87 and .18)

Essentially the entire industry for each of these three cases can be viewed as speciality goods. The industrial mold industry makes custom molds for other factories. The "special die" industry has "special" in is name. The validity of the procedure used in this section relies on the assumption that standardized goods make up the vast bulk of sales in a given industry. This assumption is not true for the above three industries and so we get anomalous results.[10]

This subsection can be summarized as follows. Proceeding under the assumption that the speciality goods sales share is small, we estimate speciality goods make up more than half of the plants in the vast majority of industries. This result emerges because of the high estimate on the speciality-good coefficient $\nu^{b,h}$ relative to standardized-good coefficient $\nu^{s,h}$ that we obtain for a typical industry h. By construction in the first stage, the BEJK model of the standardized segment fully accounts for the sales distribution. But plant counts are a different story. These depend heavily on population shares, even after accounting for the sales information through $\pi^h_{ii}(\eta^h, \Gamma^h)$.

6 Analysis of the Model

We use the model to analyze two issues. The first is the plant size/geographic concentration relationship. The second is the impact of imports.

6.1 The Plant Size/Geographic Concentration Relationship

Define a *high-concentration industry location* to be one where the sales location quotient is above 2 and in addition the location has at least 5 percent of the industry's sales. Across the 7 reclassification industries, there are 23 different high-concentration industry locations and these are listed in Table 8, sorted for each industry by descending sales quotient. The breakdown into the count and size quotient is also reported. (Recall $Q^{sales}_i = Q^{count}_i \times Q^{size}_i$.) It is clear from inspection of the data that the size margin plays an important role in contributing to how an industry expands at a location. Consider the wood furniture industry in the Highpoint area where the sales quotient is 27.7. (Note the table lists the area as

[10] It is worth noting that these three industries are also outliers in in the small plant sales share.

Greensboro/Highpoint). The breakdown is $27.7 = 4.2 \times 6.6$. Thus, average plant size in the area is 6.6 times the national average. A high contribution from the size margin holds for virtually all the 23 individual industry locations listed in Table 8. Over these 23 cases, the size quotient on average is 5.4 compared to an average count quotient of 4.3.

The last two columns contain fitted values of the size quotient for the constrained model with only a standardized segment (the BEJK model) and the full model that includes the speciality segment.[11] From the theoretical discussion in Section 2, we know that when transportation costs are not zero (and making further assumptions about the distribution of demand), the BEJK model implies that average plant size is bigger in locations that specialize in an industry. With only two exceptions, this *qualitative* pattern holds for the fitted BEJK model. However, the BEJK model fails *quantitatively* as the predicted size differences are small. The count margin is doing the main work of driving variations in Q_i^{sales}, just like in the numerical examples of the BEJK model in Table 1. When we turn to the full model and allow for the speciality segment, the predicted size quotients are much larger and close to what they are in the data (though still smaller). The average size quotient in the full model equals 4.0 compared to an average of only 1.5 in the constrained model. These differences between the full and the constrained model are driven entirely by *compositional* differences across locations between standardized and speciality goods in the full model. In fact, if we just look only at the standardized segment in the full model, the size distribution is identical to what is it in the constrained model.

Next consider the broader set of industries (the 165 remaining diffuse demand industries). We see in Table 8 at the same pattern holds with these industries as holds with the 7 reclassification industries. The mean size quotient is 5.3, so the size margin plays a big role in how locations specialize in an industry. The mean fitted value of the BEJK of 1.3 is way off. The corresponding mean in the full model is 3.3. Still too small, but much of the way there. The industry location level observations in Table 8 suggest some skewness in the distribution of the sales quotients so it is of interest go also look at the median. The same pattern holds.

6.2 Impact of the China Surge

Imports from China have surged in a number of manufacturing industries in recent years. This subsection examines how the geographic distribution of production in the United States

[11] In the constrained model, plant counts are proportational to $\pi_{ii}^h(\eta^h, \Gamma^h)$.

has shifted in response and compares the results to the prediction of the constrained model that does not allow for specialized goods (the BEJK model). If the change in imports into the United States could be attributed to regional shocks within the United States, e.g. one part of the country got a negative productivity shock, the rest of the country did not, it would not be sensible to attribute shifts across locations as an "impact of the China surge." We don't believe there is much to worry about on this score. The surge in imports has more to do with what is happening in China than regional shocks taking place here.

We classify an industry as having a China import surge if the industry experienced a 25 percentage point increase in overall imports as a share of shipments over the period 1997 to 2007 and if China's share of total imports in the industry as of 2007 exceeds 40 percent.[12] Of the 172 industries for which we have model estimates, there are 17 China surge industries. These are listed in Table 9. The overall import share of these 17 industries on average rose from .34 in 1997 to .70 in 2007. The share of imports from China increased from .26 to .61 over this same period. On average across these industries, employment declined 66 percent over the period. In the infant apparel industry, employment declined an astonishing 97 percent.

It is useful to have a comparison group of industries that (1) unlike the China surge industries have not been impacted by imports but (2) are similar to the China surge industries in being tradeable within the United States. We use food (NAICS=311) and beverage (NAICS=312) industries for this purpose. As shown in the bottom row of Table 9, imports in these industries are relatively small. On average, employment increased by 5 percent for these industries.

We take our first-stage model estimates of the pure BEJK model of the standardized sector for each of these industries and simulate the impact of large import shock on the distribution of domestic production in 2007. As discussed in the theory section, the transportation cost from China is assumed to be the same in all domestic locations. Also, the general equilibrium impact of the wage change is not taken into account. We look at the *relative* impact on the distribution of production rather than the *absolute* impact. So our answer is not very sensitive to the absolute size of the shock that is imposed. For each industry h, we set the 2007 China competitiveness index (the parameter summarizing both

[12] Imports as a share of shipments has in the denominator imports of the product plus all shipments of the product originating from domestic manufacturing plants.

productive efficiency and distance adjustments, see (11)) to

$$\lambda_{C,2007}^{h} = \frac{1}{2}\frac{\sum_{\ell=1}^{177}\left(\sum_{k=1}^{177}\gamma_k^h a_{\ell k}^h\right)}{177}.$$

The term in parenthesis is the denominator of (2), the sum of competitiveness across originations. We set $\lambda_{C,1997}^{h} = 0$. On account of this change in China's delivered productivity, China's share goes from zero to about a third of the market. We take into account that the distribution of demand differs somewhat across time periods, by using population estimates for 1997 and 2007 when calculating the quotients for each year.

For the sake of illustration, it is useful to start the discussion with what happened to the wood furniture industry in Highpoint. Recall from above that in 1997, the sales location quotient equaled $Q_{1997}^{sales} = 27.7$. Using the estimates of the BEJK model from stage 1, the China surge described above results in a predicted vale of $\hat{Q}_{2007}^{sales} = 31.0$ at the location. Highpoint's relative share increases by about 10 percent analogous to what happens in part (ii) of Proposition 2. Locations with large plants with high average productivity increase share after the shock. What actually happened is that the sales quotient in Highpoint fell to $Q_{1997}^{sales} = 12.8$. While national employment fell by 51 percent, employment in the area declined from 20,000 to 5,600, a decline of 72 percent.

For each industry, define the *primary location* to be the location with the highest 1997 location sales quotient, of those locations with at least 5 percent of 1997 sales. Table 10 presents an analysis of the 1997 primary locations for each of China surge industries. Note first that analogous to the pattern of Table 8, virtually all of these locations have very high average plant size relative to the national average, as indicated in the Q_{1997}^{size} column. Second, in every case but one, the sales quotient falls at the primary location from 1997 to 2007. Third, in every case but two, the BEJK predicted value $\hat{Q}_{2007}^{sales}$ increases at the primary location from the impact of the China surge. On average across the 17 industries, mean sales quotient in the data falls from 26.9 to 12.2 while the BEJK predicted value actually rises to 29.9. In terms of the median, in the data it falls from 19.6 to 5.3 while the BEJK predicted value rises to 23.4.

It is natural to expect regression to the mean to play some role here. A location that is number one in the 1997 rankings can only go down. To gain some perspective on the importance of regression to the mean, we look at the primary locations in the food and beverage comparison industries. These are similar to the China surge industries in that average plant size is quite high at the primary locations. Also, average the sales quotient is quite high (37.4 in the food/bev industries, 26.9 in the China surge industries). If there were a China surge for these industries, the BEJK predict sales quotients would increase at the primary locations. Mean sales quotient falls for this group of industry locations, from 37.4 to 31.6. But the decline is relatively small, only 15 percent, compared with the average 59 percent decline for the China surge industries. We get the same conclusion if we look at the median instead of the mean.

We next look at what is happening in big cities. We combine the twenty largest economic areas by population into one group and calculate the various quotients for the group as a whole. Table 11 is the analog of Table 10, with the big city aggregate serving as the location of interest instead of the primary location. The big city plants tend to be 22 percent smaller than the national average. As of 1997, sales were slightly underrepresented in big cities, $Q^{sales}_{1997} = .97$. The BEKJ model predicts a slight decrease going into 2007. The very clear pattern in Table 11 is that the sales quotient actually increased in the big city aggregate over this time period, rising to $Q^{sales}_{2007} = 1.08$. This is not a case of regression to the mean; as of 2007 when at $Q^{sales}_{1997} = .97$, it is close to the mean to begin with. (The population weighted mean across locations of Q^{sales}_{1997} equals one by construction.) The industries in the food/beverage comparison group are similar to the China surge industries as of 2007. However, there is virtually no change on average in the food/beverage comparison group over the time period.

We summarize the results as follows. In the industries impacted by a China surge, primary locations like Highpoint have substantially declined relative to the rest of the country while big city locations have gained. In the our estimated BEJK model for these industries with only standardized goods, the predictions go the other way. It is clear that the model of this paper that includes a speciality good segment is consistent with the qualitative pattern. Primary locations are the home of standardized good production and this has declined with the advent of China. Big cities are a place for speciality goods production.

References

Bernard Andrew B.; Eaton, Jonathan; Jensen, J. Bradford and Samuel Kortum. "Plants and Productivity in International Trade." *American Economic Review*, 2003, 93(4), pp. 1268-90.

Bernard, Andrew B. and J. Bradford Jensen. "Exporters, Jobs, and Wages in U.S. Manufacturing, 1976-1987." *Brookings Papers on Economic Activity: Microeconomics*, 1995, pp. 67-119.

Bernard, Andrew, Stephen Redding, and Peter Schott "Multi-product Firms and Product Switching" American Economic Review, forthcoming, 2009.

Berry, Steven T, (1994), "Estimating Discrete Choice Models of Product Differentiation," *Rand Journal of Economics* 25, No. 2, 242-262.

Dunne, Timothy; Roberts, Mark J. and Larry Samuelson. "The Growth and Failure of U.S. Manufacturing Plants." *Quarterly Journal of Economics*, November 1989, 104(4), pp. 671-98.

Eaton, Jonathan and Samuel Kortum. "Technology, Geography, and Trade." *Econometrica*, 2002, 70(5), pp. 1741-79.

Eaton, Jonathan; Kortum, Samuel and Francis Kramarz. "An Anatomy of International Trade: Evidence from French Firms." Working paper, 2005.

Ellison, G and EL Glaeser, "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, vol 105, no. 5, 889-927.

Juan Carlos Hallak, Jagadeesh Sivadasan, "Firms' Exporting Behavior under Quality Constraints, NBER Working Paper No. 14928, April 2009

Head, Keith and John Ries, (1999) "Rationalization Effects of Tariff Reductions," *Journal of International Economics* 47, vol 2, 295-320.

Hillberry, Russell and David Hummels. "Intranational Home Bias: Some Explanations." *Review of Economics and Statistics*, November 2003, 85(4), pp. 1089-92.

Hillbery, Russell and Hummels, David (2008), "Trade Responses to Geographic Frictions: A Decomposition Using MicroData," European Economic Review 52, pp. 527-550.

Holmes, Thomas J. and John J. Stevens. “Geographic Concentration and Establishment Scale.” *Review of Economics and Statistics*, November 2002, 84(4), pp.682-90.

Holmes, Thomas J. and John J. Stevens. “Geographic Concentration and Establishment Size: Analysis in an Alternative Economic Geography Model.” *Journal of Economic Geography*, June 2004a, 4(3), pp. 227-50.

Holmes, Thomas J. and John J. Stevens "Spatial Distribution of Economic Activities in North America," *Handbook on Urban and Regional Economics*, North Holland: (2004b)

Hopenhayn, Hugo. (1992), “Entry, Exit, and firm Dynamics in Long Run Equilibrium,” *Econometrica*, Vol. 60, No. 5 (Sep., 1992), pp. 1127-1150

Hopenhayn, Hugo and Richard Rogerson, “Job Turnover and Policy Evaluation: A General Equilibrium Analysis,” *The Journal of Political Economy*, Vol. 101, No. 5 (Oct., 1993)

Hsu, Wen-Tai, (2008), “Central Place Theory and Zipf’s Law” manuscript.

Hummels, David and Alexandre Skiba, “Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture.” *Journal of Political Economy* 112 (2004) 1384-1402.

Jovanovic, Boyan, (1982), “Selection and the Evolution of Industry,” *Econometrica* 50, 3, 649-670.

Atkeson, Andrew and Patrick Kehoe (2005) “Modeling and Measuging Organization Capital,” *Journal of Political Economy.*

Lucas, Robert (1978), “On the Size Distribution of Business Firms, *Bell Journal of Economics*, Vol 9, No. 2, 508-523.

Luttmer, Erzo G.J. (2007), “Selection, Growth, and the Size Distribution of Firms,” *Quarterly Journal of Economics*, Vol. 122, No. 3, 1103-1144.

Melitz, Marc. “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity.” *Econometrica*, November 2003, 71(6), pp. 1695-1725.

Office of Management and Budget (1994), Federal Register Notice, July 26, 1994, "Economic Classification Policy Committee Standard Industrial Classification Replacement.

Piore, Michael J. and Charles F. Sabel. *The Second Industrial Divide*, New York: Basic Books, 1984.

Roberts, Mark and Dylan Supina, "Output Price, Markups, and Producer Size," *European Economic Review Papers and Proceedings,* Vol. 40, Nos. 3/4 (April 1996), pp. 909-921.

Scherer, F. M. (1980),*Industrial market structure and economic performance* 2nd Edition, Chicago : Rand McNally.

Syverson.

Viner, J. (1932) "Cost Curves and Supply Curves," Zeitschrift fur Nationalokomomie, Vol 3, pp 23-46.

Table 1
Breakdown of Q_2^{sales} into Q_2^{count} and Q_2^{size} for Various Parameters
(In all cases $Q_2^{sales} = 4$)

Transportation Structure *a*	Equal Size Locations ($x_1 = x_2$)			Location 2 larger ($3x_1 = x_2$)			Location 1 larger ($x_1 = 3x_2$)		
	γ_2/γ_1	Q_2^{count}	Q_2^{size}	γ_2/γ_1	Q_2^{count}	Q_2^{size}	γ_2/γ_1	Q_2^{count}	Q_2^{size}
1.000	4.00	4.00	1.00	12.00	4.00	1.00	1.33	4.00	1.00
0.990	4.00	3.98	1.01	11.94	3.95	1.01	1.34	4.01	1.00
0.950	4.00	3.88	1.03	11.71	3.74	1.07	1.37	4.07	0.98
0.900	4.01	3.77	1.06	11.43	3.49	1.15	1.41	4.14	0.97
0.800	4.06	3.55	1.13	10.91	3.02	1.32	1.50	4.29	0.93
0.600	4.32	3.16	1.27	10.18	2.24	1.19	1.76	4.60	0.87
0.500	4.62	2.98	1.34	10.07	1.92	2.09	1.97	4.75	0.84
0.200	8.28	2.59	1.54	13.87	1.24	3.23	4.00	5.14	0.78
0.100	15.41	2.52	1.58	23.88	1.12	3.56	7.63	5.22	0.77
0.010	150.05	2.50	1.60	225.22	1.08	3.69	75.01	5.25	0.76
0.001	1501.20	2.50	1.60	2257.89	1.09	3.68	750.11	5.25	0.76

Table 2
Descriptive Statistics for the 1997 Reclassification Industries
By SIC/Retail and SIC/Man status

NAICS Industry Classification	Classification Based on SIC	Number of Plants	Mean Plant Employ.	Export Share	Mean Location Quotient
Confectionery from purchased chocolate	SIC/Retail	440	8	.00	1.01
	SIC/Man	421	70	.03	4.87
Nonchocolate confectionery	SIC/Retail	349	4	.00	1.01
	SIC/Man	276	88	.03	4.61
Curtain & drapery mills	SIC/Retail	1,085	4	.00	1.54
	SIC/Man	999	21	.03	3.22
Other apparel accessories & other apparel	SIC/Retail	724	3	.00	1.71
	SIC/Man	966	25	.03	2.67
Wood kitchen cabinet & counter top	SIC/Retail	2,055	5	.00	1.25
	SIC/Man	5,908	15	.01	2.14
Upholstered household furniture	SIC/Retail	576	5	.00	1.52
	SIC/Man	1,130	77	.03	7.20
Nonupholstered wood household furniture	SIC/Retail	815	6	.00	1.17
	SIC/Man	3,035	41	.03	4.42

Table 3
Mean Specialty Share across Sample Plants
By SIC/Retail and SIC/Man status and Within SIC/Man by Employment Size
For Five NAICS Industries Where Specialty Variable can be Defined

Industry Grouping	Classification	Number of Sample Plants	Mean Speciality Share
All (Five NAICS Industries)	SIC/Retail	160	.82
	SIC/Man	3,169	.40
	Within SIC/Man by Emp. Size		
	1-19	1,655	.57
	20-99	973	.27
	100 and above	539	.07
Confectionery (Two NAICS)	SIC/Retail	58	.82
	SIC/Man	225	.01
	Within SIC/Man by Emp. Size		
	1-19	27	.00
	20-99	96	.02
	100 and above	102	.00
Wood kitchen cabinet & counter top (One NAICS)	SIC/Retail	48	.87
	SIC/Man	1,854	.64
	Within SIC/Man by Emp. Size		
	1-19	1331	.70
	20-99	426	.56
	100 and above	97	.28
Household furniture (Two NAICS)	SIC/Retail	54	.78
	SIC/Man	1,090	.05
	Within SIC/Man by Emp. Size		
	1-19	297	.08
	20-99	451	.05
	100 and above	340	.03

Table 4
Mean Location Quotient by Plant Size for Three Groups of Industrys

Industry Grouping	Plant Size Category	Number of Establishments	Mean Location Quotient	
			Raw	NAICS Fixed Effect
Reclassification Industry Sample (7 Industries)	All	18,585	4.32	4.32
	1-19	15,687	1.36	2.46
	20-99	2,073	2.62	3.03
	100-499	690	4.18	3.97
	500+	135	6.13	5.83
Other Diffuse Demand Industries (165 Industries)	All	130,986	4.86	4.86
	1-19	91,608	1.73	3.72
	20-99	28,291	2.37	4.01
	100-499	9,533	3.75	4.60
	500+	1,554	6.68	5.41
Remaining Manufacturing Industries (301 Industries)	All	211,945	5.13	5.13
	1-19	134,044	1.86	3.67
	20-99	54,705	2.79	4.14
	100-499	20,101	4.38	4.91
	500+	3,095	6.92	5.82

Table 5
Model Estimates from Stage 1

	Parameter Estimates (s.e. in parenthesis)		Implied value of a(dist) by dist			Number of CFS Shipments Used for Estimate
	η_1	η_2	100 miles	500 miles	1000 miles	
Baseline 1997 NAICS Estimates using $\overline{dist} = 100$						
Selected Industries by percentile of implied value of a(100)						
Ready-mix concrete (Minimum)	.0523 (.0004)	0*	.01	.00	.00	37,875
Creamery butter (25^{th} percentile)	.0030 (.0003)	-8.1E-07 (1.1E-07)	.74	.27	.11	586
Wood television, cabinets (50^{th} percentile)	.0019 (.0003)	-6.4E-07 (1.6E-07)	.83	.45	.28	546
Surgical appliance & supplies (75^{th} percentile)	.0008 (.0001)	-1.8E-07 (0.2E-07)	.92	.69	.52	11,670
Other hosiery & sock mills (Maximum)	0**	0**	1.00**	1.00**	1.00**	2,967
Baseline Estimates (Mean over 172 Industries)	.0030 (.0002)	-5.0E-07 (.0.7E-07)	.80	.48	.33	3,968
Alternative Estimates (Means across industries)						
1992 SIC (175 Industries) with $\overline{dist} = 100$	.0030 (.0001)	-5.9E-07 (0.4E-07)	.79	.44	.30	8,500
1997 NAICS (172 Industries) with $\overline{dist} = 0$			.71	.30	.16	3,968
1997 NAICS (172 Industries) with $\overline{dist} = 200$			.85	.57	.42	3,968

*The constraint η_2=0 is imposed for this industry.
**This estimate is at the constraint that a(dist)≤1

Table 6
The Model's Goodness of Fit of the Distance Distribution of Shipments
Conditioned upon Shipments being at least 100 miles

Statistic Reported	Category 1 100≤distance<500	Category 2 500≤distance<1000	Category 3 1000≤distance
Mean s_c^h across industries (data)	.44	.30	.25
Mean $\hat{s}_c^h$ across industries (model)	.38	.33	.30
Regression of s_c^h (data) on $\hat{s}_c^h$ (model)			
Intercept	.05 (.02)	-.03 (.02)	-.03 (.01)
Slope	1.04 (.04)	.99 (.07)	.96 (.03)
R^2	.81	.55	.83
Number of Observations	167	167	167

Table 7
Second Stage Estimates of the Plant Count Parameters and Related Model and Data Statistics

	Regression Results (s.e. in parentheses)				**Model**	**Data** Small Plants (Employment<20)	
	Constant λ	Slope (spec) ν^b	Slope (stan) ν^s	R^2	Implied Specialty plant share	Share of plant counts	Share of sales
Reclassification Industries							
Confectionery from purchased chocolate	.4	614.0	104.5	.69	.79	.77	.04
	(.2)	(52.5)	(17.4)				
Nonchocolate confectionery	.1	481.5	63.4	.79	.79	.77	.03
	(.1)	(31.5)	(17.4)				
Curtain & drapery mills	-.9	2162.0	19.2	.92	.96	.91	.24
	(.3)	(59.8)	(6.9)				
Other apparel accessories & other apparel mfg	-2.3	1480.8	145.8	.86	.67	.84	.17
	(.5)	(136.7)	(15.5)				
Wood kitchen cabinet & counter top mfg	1.0	6059.4	130.8	.88	.79	.89	.28
	(1.2)	(238.0)	(15.1)				
Upholstered household furniture mfg	-1.3	905.0	240.5	.89	.46	.69	.05
	(.5)	(92.4)	(7.4)				
Nonupholstered wood household furniture mfg	-1.1	3195.4	228.1	.86	.78	.81	.08
	(.7)	(139.8)	(19.5)				
Means of 7 Reclassification Industries	-.6	2128.3	133.2	.84	.75	.81	.13
	(.5)	(107.2)	(12.8)				
Means of 165 Remaining Diffuse Demand Industries	-.3	634.7	75.2	.71	.65	.57	.08
	(.2)	(41.6)	(5.9)				

Table 8 Sales, Count and Size Quotients in Data, Size Quotients for Both Models In High Concentration Industry Locations

Industry (of 7 reclassification industries)	Location	Mean sales share	Data Q^{Sales}	Data Q^{Count}	Data Q^{Size}	BEKJ Only Model Q^{Siz}	Full Model Q^{Size}
Confectionery from purchased chocolat	Harrisburg- PA	.07	9.2	1.4	6.4	1.2	3.8
	Nashville, TN	.06	6.6	.8	7.9	1.2	3.4
	Chicago	.15	4.1	1.2	3.6	1.2	2.8
	Philadelphia	.08	3.3	2.1	1.6	1.1	2.4
	San Francisco, CA	.08	2.3	1.5	1.6	.6	1.4
Nonchocolate confectionery mfg	Grand Rapids- MI	.07	11.2	1.2	9.2	1.4	4.5
	Chicago, IL	.24	6.8	1.5	4.6	1.4	3.8
	Atlanta- GA	.07	3.5	.8	4.5	1.2	2.5
Curtain & drapery mills	San Antonio, TX	.07	1.3	.6	16.8	.7	6.8
	Raleigh-Durham NC	.09	1.1	1.1	8.9	1.6	8.4
	Charlotte- NC	.06	7.6	1.7	4.5	1.8	6.7
	Boston- MA	.07	2.5	1.4	1.8	1.1	2.3
Other apparel	New York- NY	.28	3.5	2.6	1.4	1.9	2.8
	Los Angeles-, CA	.16	2.5	2.1	1.2	1.0	1.7
Wood kitchen cabinet & counter top mfg	Harrisburg- PA	.05	7.4	1.7	4.4	4.8	6.6
	Dallas-Fort Worth, TX	.05	2.4	1.0	2.4	1.2	2.0
Upholstered household furniture mfg	Tupelo, MS	.21	107.4	43.4	2.5	1.6	2.8
	Charlotte- NC	.19	23.2	11.3	2.1	1.8	3.1
	Knoxville- TN	.09	22.2	2.6	8.6	1.7	3.0
	Greensboro,HighPoint,	.12	19.2	11.0	1.8	1.8	3.1
Nonupholstered wood household furniture	Greensboro,HighPoint,	.17	27.7	4.2	6.6	2.0	7.2
	Charlotte- NC	.13	15.5	2.9	5.4	1.9	6.1
	Toledo-Fremont, OH	.05	13.5	.8	17.5	1.5	5.0
Summary Statistics							
7 Reclassification Industries	N= 23 Industry Locations						
	Mean	.11	14.0	4.3	5.4	1.5	4.0
	Median	.08	7.6	1.5	4.5	1.4	3.1
165 Remaining Diffuse Demand Industries	N= 566 Industry Locations						
	Mean	.11	18.3	5.9	5.3	1.3	3.3
	Median	.09	9.3	2.9	2.6	1.2	2.5

Table 9
List of China Surge Industries and Some Descriptive Statistics

Industry	Import Share of Shipments 1997	Import Share of Shipments 2007	China Share of Imports 1997	China Share of Imports 2007	Employment Growth 1997-2007
Curtain & drapery mills	.08	.56	.38	.65	-.47
Other household textile prod mill	.22	.68	.25	.49	-.51
Women's & girls' cut & sew dress	.29	.67	.21	.55	-.71
Women's & girls' cut & sew suit,	.48	.92	.19	.49	-.91
Infants' cut & sew apparel mfg	.60	.99	.08	.62	-.97
Hat, cap, & millinery mfg	.44	.80	.26	.67	-.74
Glove & mitten mfg	.58	.88	.50	.63	-.78
Men's & boys' neckwear mfg	.25	.56	.02	.59	-.67
Other apparel accessories	.39	.80	.35	.64	-.75
Blankbook, looseleaf binder,	.18	.47	.43	.52	-.51
Power-driven handtool mfg	.28	.56	.18	.46	-.56
Electronic computer mfg	.12	.49	.00	.56	-.68
Electric housewares & fan mfg	.52	.78	.48	.76	-.54
Wood household furniture mfg	.29	.62	.18	.46	-.51
Metal household furniture mfg	.29	.55	.37	.85	-.48
Silverware & plated ware mfg	.44	.91	.31	.73	-.82
Costume jewelry & novelty mfg	.31	.68	.31	.67	-.63
Mean of China Surge Industries (N=17)	.34	.70	.26	.61	-.66
Mean of Food/Bev Comparison Group Industries (N=35)	.08	.11	.02	.07	.05

Table 10: Analysis of 1997 Primary Industry Location
Data and Fitted Values in BEJK Model after China Surge

Industry	Data: Q_{1997}^{size}	Data: Q_{1997}^{sales}	Data: Q_{2007}^{sales}	BEJK Predicted Value After China Surge $\hat{Q}_{2007}^{sales}$
Curtain & drapery mills	16.8	10.3	5.3	7.9
Other household textile prod mill	3.6	19.2	9.2	20.6
Women's & girls' cut & sew dress	1.0	4.7	4.1	4.8
Women's & girls' cut & sew suit,	3.8	4.7	0.9	4.9
Infants' cut & sew apparel mfg	3.8	68.2	0.0	76.9
Hat, cap, & millinery mfg	39.4	71.3	11.5	79.8
Glove & mitten mfg	1.7	32.9	67.9	36.9
Men's & boys' neckwear mfg	3.5	17.7	0.2	21.6
Other apparel accessories	1.4	3.5	2.6	3.8
Blankbook, looseleaf binder,	2.5	6.8	5.3	5.4
Power-driven handtool mfg	6.5	54.4	47.1	59.6
Electronic computer mfg	17.0	43.0	0.5	47.7
Electric housewares & fan mfg	4.0	23.6	23.5	25.8
Wood household furniture mfg	6.6	27.7	12.8	31.0
Metal household furniture mfg	4.5	7.6	1.0	7.9
Silverware & plated ware mfg	17.3	42.4	0.6	50.0
Costume jewelry & novelty mfg	1.6	19.6	14.4	23.4
Summary Statistics				
China Surge Industries (N=17)				
Mean	7.9	26.9	12.2	29.9
Median	3.8	19.6	5.3	23.4
Food/Bev Comparison Group (N=32)				
Mean	6.2	37.4	31.6	38.4
Median	3.3	21.3	14.2	21.6

Notes: 1997 primary industry location is location with highest Q_{1997}^{sales} of those locations with at least 5 percent of industry sales.

Table 11: Analysis of Ten Largest Economic Areas by Population
Data and Fitted Values in BEJK Model after China Surge

Industry	Data Q^{size}_{1997}	Data Q^{sales}_{1997}	Data Q^{sales}_{2007}	BEJK Predicted Value After China Surge $\hat{Q}^{sales}_{2007}$
Curtain & drapery mills	0.79	0.85	1.06	0.81
Other household textile prod mill	0.55	0.63	0.73	0.56
Women's & girls' cut & sew dress	0.97	1.66	1.72	1.67
Women's & girls' cut & sew suit,	0.80	1.28	1.59	1.28
Infants' cut & sew apparel mfg	0.76	0.65	0.88	0.66
Hat, cap, & millinery mfg	0.67	0.87	1.04	0.86
Glove & mitten mfg	1.04	0.45	0.48	0.41
Men's & boys' neckwear mfg	0.91	1.51	1.72	1.51
Other apparel accessories	1.02	1.39	1.31	1.40
Blankbook, looseleaf binder,	0.93	1.24	1.19	1.24
Power-driven handtool mfg	0.54	0.58	0.60	0.58
Electronic computer mfg	0.60	0.79	0.75	0.81
Electric housewares & fan mfg	0.43	0.42	0.71	0.40
Wood household furniture mfg	0.45	0.44	0.57	0.35
Metal household furniture mfg	0.91	1.03	1.09	1.01
Silverware & plated ware mfg	0.78	0.97	1.38	0.97
Costume jewelry & novelty mfg	1.12	1.73	1.57	1.78
Summary Statistics				
China Surge Industries (N=17)				
Mean	.78	.97	1.08	.96
Median	.79	.87	1.06	.86
Food/Bev Comparison Group (N=35)				
Mean	.94	.91	.90	.92
Median	.93	.89	.90	.92

Notes: The ten largest economic areas by population are grouped into one aggregate big city area. The statistics reported above are calculated for the big city area being treated as a single location.

Figure 1
A Model Statistic and a Related Data Statistic
7 Reclassification Industries

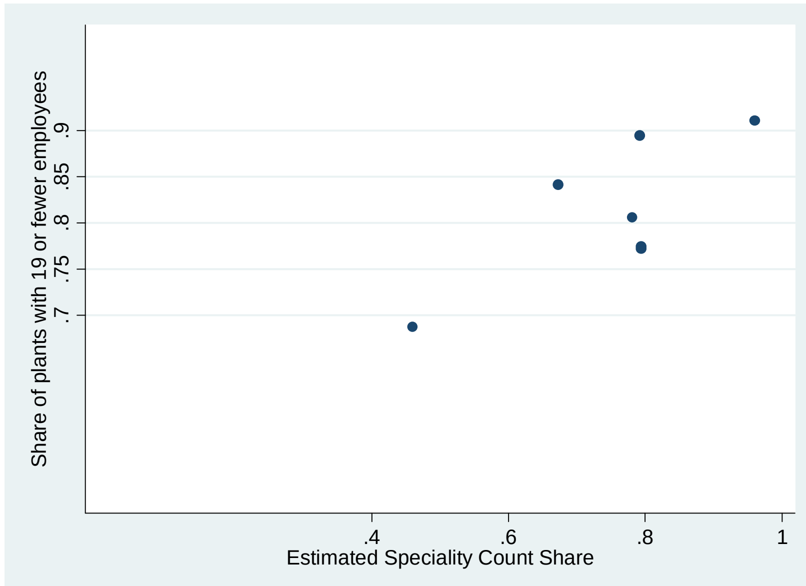

Figure 2
A Model Statistic and a Related Data Statistic
165 Remaining Diffuse Demand Industries