Robust Structural Equations for Designing and Monitoring Strategic International Facility Networks

Panos Kouvelis
Olin School of Business, Washington University, St. Louis, Missouri 63130, USA, kouvelis@olin.wustl.edu

Charles L. Munson
College of Business, Washington State University, Pullman, Washington 99164-4746, USA, munson@wsu.edu

Shilei Yang
School of Business Administration, Southwestern University of Finance and Economics, Chengdu, Sichuan 610074, China, syang@swufe.edu.cn

Using predictive global sensitivity analysis, we develop a structural equations model to abstract from the details of a large-scale mixed integer program (MIP) to capture essential design trade-offs of global manufacturing and distribution networks. We provide a conceptual framework that describes a firm’s network structure along three dimensions: market focus, plant focus, and network dispersion. Normalized dependent variables are specified that act as proxies for a company’s placement into our conceptual network classification via the calculation of just a few key independent variables. We provide robust equation sets for eight cost structure clusters. Many different product types could be classified into one of these groups, which would allow managers to use the equations directly without needing to run the MIP for themselves. Our numerical tests suggest that the formulas representing the network structure drivers—economies of scale, complexity costs, transportation costs, and tariffs—may be sufficient for managers to design their strategic network structures, and perhaps more importantly, to monitor them over time to detect potential need for adjustment.

Key words: global network design; plant location; structural equations modeling; supply chain management

History: Received: March 2010; Accepted: May 2012 by Aleda Roth, after 2 revisions.

1. Introduction

Firms that produce, distribute, and sell products worldwide face the particularly challenging task of designing appropriate facility networks. They must examine numerous factors, including product-mix decisions based on demand projections for different markets, transportation costs, and production costs. With that information in hand, companies have to decide, among other things, where to locate factories, how to allocate production activities to the various facilities of the network, and how to manage the distribution of products (e.g., where to locate distribution facilities).

Such a task can seem overwhelming for the typical management team. Although various mathematical models have been developed for global facility location and distribution problems, most are challenging and time consuming to formulate, and solution times may be formidable. In this article, we introduce the technique of predictive global sensitivity analysis to potentially remove tedious modeling requirements from managers and provide them with a set of “black-box” structural equations that are relatively easy to calculate and appear to provide good direction to indicate the type of global network structure for firms to utilize. We claim that these equations are robust because they have been developed to cover eight different cost structure clusters that are mutually exclusive and cover the entire range of parameters described by our modeling environment. We provide a mechanism for managers to quickly determine which group their product falls under—hence which set of structural equations to use. Numerical validation tests described in section 5.1 show promising results.

The main contributions of this article are:

1. A conceptual framework classifying strategic global facility network structures along three dimensions that can be operationalized as a managerial tool to develop and monitor global networks;
2. A proposed approach that we label “predictive global sensitivity analysis” for the development of structural equations that allows managers to classify their network structures via relatively simple calculations of primary levers, without the need to explicitly formulate and solve an MIP; and
3 An illustration of how to create sets of mutually exclusive robust equations that cover the entire range of parameters described by our modeling environment and how to develop a mechanism for determining which group a given product falls into. Such an exercise would enable a manager of any product fitting the modeling environment to go directly to the structural equations without having to develop those equations herself.

1.1. A Conceptual Framework to Classify Global Network Structures

The three most applicable streams of research related to characterizing global network structures address international operations, focused factories, and facility location. Within the international operations literature, numerous factors, many of which are more qualitative than quantitative, are often cited as contributing to the global facility location decision (Werner et al. 2009). In terms of number of production facilities to locate around the world, a basic trade-off often involves the concept of whether or not firms should produce inside their major markets (Ferdows 1997a, Heizer and Render 2011). We use the term market focus to describe this concept. Firms that produce in a given marketing and sales region can avoid tariffs and take advantage of lower transportation costs. On the other hand, low-cost plants, often located in low-cost countries, can be built to satisfy worldwide demand when those cost savings outweigh potentially higher tariffs and transportation costs (Cohen and Lee 1989, Goetschalckx et al. 2002, Rudberg and West 2008).

Skinner (1974) coined the term “focused factory” to describe a plant that focuses on only one specific manufacturing task. Manufacturing firms using a focused factory approach would have many more plants in the network than ones that combined subassembly manufacturing tasks. Such diseconomies can be quantified in the form of the complexity costs of not being focused (Cattani et al. 2010). Alternatively, economies of scale/scope benefits such as shared facility expenses, managerial expertise, administrative costs, and purchasing discounts might encourage firms to produce multiple subassemblies at one location.

At the core of the facility location-allocation literature, a basic trade-off typically involves the concept of “many vs. few” (Simchi-Levi et al. 2008). In other words, high fixed costs or economies of scale may encourage firms to have fewer plants producing more units each, while high transportation costs or service level considerations may encourage firms to have many plants closer to customers that produce fewer units each (Klose and Drexl 2005, Şahin and Sural 2007). We use the term network dispersion to describe this concept.

With these considerations in mind, we propose a conceptual framework that classifies global network structures along three dimensions (Figure 1), which, when taken together, would indicate the appropriate number of production facilities worldwide. All of our succeeding model development is designed to suggest to firms where they fit along this three-dimensional classification. (Note that our definitions and the modeling environment described in section 3.1 focus on the production network of subassemblies used to produce a single final product; however, the conceptual framework itself could just as easily apply to the production network of multiple final products.)

The first dimension, market focus, refers to the degree to which each marketing region produces subassemblies needed for that region. A global strategy generally has plants producing large fractions of worldwide demand, while a regional strategy generally segments production by region. Second, plant focus refers to the degree to which different subassemblies are produced at the same facility. A dedicated strategy generally separates subassemblies into different plants, while a flexible strategy combines subassembly production at the same facilities. Third, network dispersion refers to the number of different plants producing the same subassembly per marketing region that has any production at all. This dimension accounts for the number of redundant plants for each subassembly, after controlling for the degree of market focus and plant focus. If subassembly production is occurring in a region, then a concentrated strategy would produce any given subassembly at no...
more than one facility (whether as part of a dedicated or flexible plant focus strategy), whereas a scattered strategy would produce the subassembly at multiple plants, for example, at a plant in every country in the region.

Theoretically, a firm’s network structure could fall anywhere along the axes of Figure 1. Figure 2 provides a comparative illustration of the two extreme conditions, as well as a mixed condition on each dimension (Figure 2b), based on the same example set of marketing regions, countries, subassemblies, and distribution centers (DCs). At one extreme, a global-flexible-concentrated strategy (Figure 2c) would have all subassemblies for the entire world’s demand produced at one facility. At the other extreme, a regional-dedicated-scattered strategy (Figure 2a) would have each marketing region producing to meet its own demand at multiple subassembly-specific facilities.

Based on the above discussion, the literature suggests that economies of scale, complexity costs of producing different subassemblies at the same plant, transportation costs, and tariffs are the primary drivers in determining a company’s network structure according to our classification. (Of course, many other factors can affect network structure, some of which are included in our MIP, but for aggregation purposes we expect these four factors to be among the most important.) Our research enhances the level of understanding by first developing effective aggregate ways to measure “economies of scale,” “complexity costs,” and relevant “transportation” and “tariff” costs, and then explicitly describing via empirically derived equations the causal relationship of these measures with appropriately defined measures of “market focus,” “plant focus,” and “network dispersion.” This allows us to generate a set of tools, a set of structural equations, that could be adopted by the global manager to construct and monitor the firm’s global facility network.

2. Related Literature

Other authors have emphasized certain other factors when classifying global network structures. The typical textbook approaches employ an “orientation of facility networks” concept (e.g., Dornier et al. 1998), where firms are assumed to configure supply chain networks following one of three strategies: market focus, product family focus, or process focus. Ferdows (1997b) identifies six foreign factory types based on reasons for the factory’s location and the site’s competence level. The types are “offshore” (access to low cost production/low competence), “source” (access to low cost production/high competence), “outpost” (access to skills and knowledge/low competence), “lead” (access to skills and knowledge/...
help select all of the individual recommended changes.

Preliminary results from this work are reported in Kouvelis and Munson (2004). In that article, the authors show how to develop structural equations for one company at a time. They also perform extensive numerical tests to illustrate the importance of incorporating taxes, depreciation, and financing in a plant-location model. While the modeling approach in Kouvelis and Munson (2004) yields impressive validation results, it suffers from a lack of robustness. Specifically, the equations provided might only apply for the particular sample of products used to build the equations in the first place.

This article extends the work of Kouvelis and Munson (2004) in several ways, including the following. First, this article proposes an “extra layer” of modeling for researchers interested in applying Wagner’s (1995) global sensitivity analysis to predict complications in operations research. Specifically, we illustrate how to create sets of more robust equations that should work well for any particular product once it is appropriately classified into its designated set. Second, the validation methods used are completely different and, we think, significantly improved. Third, we provide a numerical example that takes the reader step-by-step through the process of selecting the correct cluster, computing the independent variable values, estimating the dependent variable values via the structural equations, and via sensitivity analysis showing how the structural equations might be used to monitor and perhaps alter the global facility network over time.

3. Structural Equations Model

3.1. Problem Statement

A firm considers the issue of designing its global network of plants and DCs. The firm plans to produce a new product with sales expectations in many countries. The product is composed of many different subassemblies, but requires only one unit per subassembly (i.e., a two-level product tree, with level 0 the final product and level 1 all its subassemblies). Subassemblies can be produced in some combination of facilities within the firm’s market areas. To assemble the final product, DCs can be located in various countries. The firm wants to develop a network of plants and DCs that maximizes its discounted after-tax profit over the planning horizon for this product.

3.2. MIP Model for Facility Location

An MIP model such as that presented in Appendix A can be used to solve the global plant location problem just described. We chose this particular version of an MIP model to address the areas of emphasis in this
We now shift our emphasis toward identifying a few such aggregate measures can be used to effectively information from the MIP model. We illustrate how subsequently determining the cost for land and buildings. mine the overall size of the production facility, conse-
demand allocation placed upon it. Then, the size of all production line costs consist of the equipment to run a plant that is twice as big will have certain higher proportion thus approximates the idea, for example, that each potential facility may not be known. The linear portion becomes “fixed” after the location-
allocation decisions are made, but it is included in the model at the beginning because until those location-
allocation decisions are known, the relative size of each potential facility may not be known. The linear portion thus approximates the idea, for example, that a plant that is twice as big will have certain higher fixed costs (e.g., electricity) and investment costs (e.g., construction materials). In cases where potential facility sizes and/or production line sizes are known up front and not determined by the location-
allocation decision, the “linear fixed” and “linear investment” portions of the model could be ignored. Furthermore, for the plants, we have split investment costs into the costs for production lines and the costs for the land and buildings around those lines. The production line costs consist of the equipment to run the line, and the size of the line may depend upon the demand allocation placed upon it. Then, the size of all of the lines produced at the same facility would determine the overall size of the production facility, consequently determining the cost for land and buildings.

3.3. Predictive Global Sensitivity Analysis

We now shift our emphasis toward identifying a few key measures that summarize the relevant input information from the MIP model. We illustrate how such aggregate measures can be used to effectively predict the nature of the optimal solution to the model (i.e., how the optimal network structure superimposes against our proposed conceptual framework) without having to solve it. In accomplishing this task, we follow Wagner’s (1995) “approximation approach” for global sensitivity analysis. Note that the predictive modeling approach described here could theoretically be applied to almost any type of MIP model—the approach itself is not dependent upon the specific MIP in Appendix A.

To use the approximation approach to global sensitivity analysis, one first starts with a deterministic model for the problem at hand. The model is run multiple times for different values of the input parameters to generate a dataset consisting of combinations of input parameters with associated optimal decision variable values. When regressions are run on the dataset, the “dependent variables” are composed of combinations of select model decision variables that the modeler deems of interest. The “independent variables” are generally aggregations of select input parameters to the model. The modeler uses judgment to create the set of potentially important independent variables, which may be any of the model coefficients themselves or quite possibly summary functions of model coefficients.

To determine the most important independent variables out of the set of potential ones, “one-at-a-time regressions” are then performed for each potential independent variable in the set by attempting a polynomial fit that includes terms \( x^4 \), where \( A \in \{ -6, -5, \ldots, -1, 1, \ldots, 5, 6 \} \). This exercise gauges the effect of each potential individual independent variable on each of the dependent variables. The independent variables exhibiting the largest adjusted \( R^2 \) values would typically be deemed to have the most influence.

The next step is to take the few most important variables and combine them into one regression. Interaction effects and nonlinearities are approximated by combining the independent variables. For example, when combining three variables, we included the 36 terms indicated above, plus 24 terms of the form \( x^{A_1}y^{A_2}z^{A_3} \), where \( A_1 \in \{ -2, -1, 1, 2 \} \) and \( A_2, A_3 \in \{ \pm A_1 \} \). Stepwise regression is then applied to all of the terms (we used SPSS® for Windows®) keeping the default alpha values of 0.05 to enter and 0.10 to remove. In other words, at each step, each remaining independent variable is separately added to the model, and the variable providing the best F-statistic is kept, as long as the statistic is significant at the \( x \) level of 0.05. Then a recheck is performed on the variables currently in the model, and the ones whose F-statistic is no longer significant at the \( x \) level of 0.10 are removed. Once completed, the stepwise regression results constitute our structural equations, which hopefully exhibit high adjusted \( R^2 \) values.

The Wagner (1995) article focuses on ways to identify variables that have a strong influence on the model results. He mentions almost in passing that the coefficients resulting from the stepwise regression
represent a “good numerical approximation” for the objective function. We actually view these techniques as ones that not only can find good explanatory variables, but also can actually predict objective function values or other dependent variable values. We coin such an extension to Wagner’s techniques as “predictive global sensitivity analysis.”

In our article, key inputs are incorporated into just four primary levers. The model outputs as incorporated into proxy measures of the structural classifications of Figure 1 are then regressed against the primary input levers. The regression results create our structural equations.

3.4. Framework Measures (Independent Variables)

As described in section 1, the literature has led us to propose four “framework measures” that we believe will have the largest impact on the type of facility network appropriate for a company: economies of scale, complexity, transportation with tariffs, and transportation without tariffs.

The framework measure calculations are shown in Equations (1)–(4) below, with associated notation provided in Table 1. Via the use of cost ratios, all four measures are dimensionless in nature. Each of our measures incorporates system-wide averages of a subset of the inputs required for the full MIP. The formulae strike a balance between “richness” and “ease of use” while attempting to capture the “essence” of the four frameworks. A modeler could certainly construct measures that contain much more detail, but the whole point of utilizing a structural equations approach like ours is to be able to compute a few summary measures quickly and then evaluate the equation results right away. We should also point out that alternative formulae might work equally well, as long as they include the important driving cost elements for that particular independent variable.

Table 1: Notation for the Independent Variables*

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T )</td>
<td>Time horizon</td>
</tr>
<tr>
<td>( F(t) )</td>
<td>Annual fixed base cost (linear cost per square foot) of operating a subassembly plant</td>
</tr>
<tr>
<td>( A )</td>
<td>Base investment cost (investment cost per square foot) to build a subassembly plant</td>
</tr>
<tr>
<td>( SUME )</td>
<td>Sum across all subassemblies of base (unit) investment cost to build a production line</td>
</tr>
<tr>
<td>( SUMV )</td>
<td>Sum across all subassemblies of unit variable production cost</td>
</tr>
<tr>
<td>( G )</td>
<td>Average space requirement per subassembly unit</td>
</tr>
<tr>
<td>( WORLD )</td>
<td>Average annual worldwide demand</td>
</tr>
<tr>
<td>( B_2 )</td>
<td>Annual fixed cost of producing a second subassembly in the same plant</td>
</tr>
<tr>
<td>( CNTY )</td>
<td>Number of potential production countries</td>
</tr>
<tr>
<td>( MILE )</td>
<td>Transportation cost/mile for the final product (excluding any flat-rate shipping costs)</td>
</tr>
<tr>
<td>( AVGMI )</td>
<td>Average shipping distance between countries</td>
</tr>
<tr>
<td>( TARIFF )</td>
<td>Average tariff for shipping the final product outside of the marketing region</td>
</tr>
<tr>
<td>( v )</td>
<td>Unit variable assembly cost at a distribution center</td>
</tr>
</tbody>
</table>

*All costs and prices are based on a system-wide average.

are particularly high (relative to profit margin per unit), the EOS measure implies that fewer, bigger plants may be required, and vice versa. So a high value of EOS would be expected to increase the likelihood of a more global market focus, a more flexible plant focus, and a more concentrated network dispersion strategy. We further note that a second form of economies of scale could exist, that is, a decreasing marginal cost per unit as volume increases. If a modeler adjusts the MIP accordingly, then the EOS summary measure could easily be modified as well, where the numerator would still represent the break-even volume.

The complexity measure, COM, calculates a ratio of the costs of producing two subassemblies at one plant vs. two subassemblies at two different plants. The measure is designed to indicate whether or not it may be cost effective to combine different subassemblies in the same facility by focusing on the relative size of the \( B_2 \) complexity cost compared to the base fixed costs of operating and building a plant. The measure assumes

\[
\text{Economies of Scale : } EOS = \frac{[TF + A + SUME]/[P - SUMV - G(Tf + a) - SUMe]}{\text{WORLD}},
\]

\[
\text{Complexity : } COM = \frac{T(F + B_2) + A + 2(WORLD/CNTY)G(Tf + a)}{2 \times [TF + A + (WORLD/CNTY)G(Tf + a)]},
\]

\[
\text{Transportation with Tariffs : } \text{TRANS1} = \frac{MILE \times AVGMI + TARIFF}{P - SUMV - v},
\]

\[
\text{Transportation without Tariffs : } \text{TRANS2} = \frac{MILE \times AVGMI}{P - SUMV - v}.
\]
that the size of a combined plant would be roughly twice the size of each plant that would produce a single subassembly. However, the base operating and investment costs in the numerator equal half of those in the denominator (by operating one plant instead of two). Finally, the numerator is increased by the complexity cost of producing a second subassembly in the same facility. Note that our COM formula only looks at the second subassembly produced at a plant and ignores any potential additional ones. We do this primarily to retain the same formula across products and industries, no matter how many subassemblies exist. The COM measure is certainly bounded below by 0.5 (as $B_2$ approaches 0), and it has no upper bound. A COM value less than 1 would generally indicate that economies of scope exist, while a value exceeding 1 would indicate diseconomies of scope. Thus, a higher value of COM would increase the likelihood of a more dedicated plant focus.

The transportation with tariffs measure, TRANS1, calculates the distance-related transportation costs plus tariffs, expressed as a percentage of gross margin. The TRANS1 measure is used primarily for the market focus strategy, because the combination of transportation costs and tariffs represents the overall cost of shipping outside of a marketing region and hence should strongly influence the likelihood of either a regional or global strategy. Finally, the transportation without tariffs measure, TRANS2, is the same as TRANS1, excluding the tariffs. This measure is primarily designed to help indicate the degree of network dispersion (tariffs do not apply within a marketing region), where a high value of TRANS2 should lead to more of a scattered strategy.

3.5. Proxies for the Global Network Structure Dimensions (Dependent Variables)

Equations (5)–(7) below display the calculations for proxies of the three network structure dimensions (the dependent variables in the structural equations model), with associated notation provided in Table 2.

$$\text{Market Focus:} \quad \text{GLOBAL} = \sum_{i=1}^{\text{REG}} \left( \frac{1}{I} \sum_{t=1}^{N} \left( 1 - \min \left\{ 1, \sum_{m \in \text{REG}} \sum_{q=1}^{I} \sum_{k=1}^{T} X_{mqkt} / \sum_{m \in \text{REG}} \sum_{q=1}^{I} \sum_{k=1}^{T} X_{mqkt} \right\} \right) \right),$$

$$\text{Plant Focus:} \quad \text{FLEX} = \sum_{m=1}^{\text{MARK}} \left( \frac{1}{I} \sum_{q=1}^{N} Y_{mq} / \sum_{m=1}^{\text{MARK}} \sum_{q=1}^{N} Y_{mq} \right) - 1,$$

$$\text{Network Dispersion:} \quad \text{SCATTER} = \sum_{i=1}^{\text{REG}} \min \left\{ 1, \sum_{m \in \text{REG}} \sum_{q=1}^{I} Z_{mq} / \max \left\{ 1, \sum_{m \in \text{REG}} \sum_{q=1}^{I} Z_{mq} \right\} \right\},$$

Each measure is normalized to a range between 0 and 1. For example, a score of 0 on the market focus measure, GLOBAL, represents a pure regional strategy, 1 represents a pure global strategy, and 0.5 represents a “partial-global, partial-regional” strategy. The measures are computed from the output of the MIP model using certain optimal decision variable values.

The market focus measure, GLOBAL, calculates the percentage of subassemblies imported from other regions, averaged over each region. A high value implies more of a global strategy. We subtract 1 from MARK in the denominator to normalize the measure. The plant focus measure, FLEX, is based on the average number of subassemblies per plant. A high value implies more of a flexible strategy. We subtract 1 from the numerator and denominator to normalize the measure. The network dispersion measure, SCATTER, calculates the average number of plants per subassembly, as a percentage of the number of possible plant locations per region, averaged over each region that has any subassembly production at all (to account for both a global-scattered strategy and a regional-scattered strategy). A high value of SCATTER implies more of a scattered strategy. As with the other two measures, we normalize SCATTER by subtracting 1 (from $(M[\text{REG}])$).

4. Robust Equations for Cost Structure Clusters

4.1. Background

Given adequate time and resources, we recommend developing product-specific structural equations stemming from random data generated from that
product’s specific costs. However, knowing that managers might lack sufficient time or expertise to implement that approach themselves on a regular basis, we illustrate how to create a set of robust equations, divided into general clusters that can be used directly by managers and that are designed to handle products from a wide variety of industries that fit within our modeling environment. A simple decision tree directs managers to the appropriate cluster.

4.2. Cluster Creation

As with any clustering exercise, choosing an appropriate number of clusters represents a trade-off between dealing with a practical number of different clusters while having the resulting models for each cluster be “good enough.” In our case, cluster analysis led us to develop eight clusters. Clearly, further refinement into more clusters would create an even better match for all products in the real world. Nevertheless, our experiments suggest that these eight particular clusters represent a manageable balance that should adequately cover many products in industry. In practice, if the general equation results seem unrealistic for any particular product, then managers should certainly consider creating their own equations via the approach described in section 3.

To create clusters, we selected 80 different products with a wide range of cost structures. Variable production costs ranged from about $1 to $20,000 per unit, which would cover many consumer and business-to-business products. Products whose costs exceed this range or have many major subassemblies would likely not fit well within the assumed operating environment of our models. Through a variety of sources, we obtained rough cost estimates for these products. We should emphasize at this point that precise cost estimates are not needed to develop the structural equations because the technique creates hundreds of random deviations from the baseline costs that cover a very wide range. The whole point is to generate many different sets of independent variable/dependent variable pairs so that the structural equations can estimate how the dependent variables change when the independent variables change.

We broadly based the clustering on variable costs (for subassembly production and assembly) along with the number of subassemblies (dividing “simpler” vs. more “complex” products). After grouping in that way, the structural equations themselves would then account for differences within the group among fixed costs, investment costs, complexity costs, transportation costs, and tariffs. Two variable cost factors, $R_1$ (reflecting a ratio of subassembly production cost to assembly cost) and $R_2$ (reflecting a ratio of subassembly production cost to flat-rate transportation cost), emerged as good measures to help break our 80 products into clusters.

Cluster analysis divided our original 80 products into eight clusters of 10 products each according to the breakpoints shown in the decision tree in Figure 3. These particular breakpoints were data-driven and based on the 80 product types with which we began. It is important to recognize at this stage that these clusters are not necessarily designed to include

**Figure 3 Decision Tree to Classify a Product into One of Eight Cost Structure Clusters (CLU)**

Notes: $R_1 = \text{SUMV} \div v$; $R_2 = \text{SUMV} \div \text{FRT}$; $\text{FRT} = \text{flat-rate portion of the per-unit transportation cost}$; and $I = \text{number of subassemblies}$.
any specific industry or even product. The clusters are based on relative cost structures and number of subassemblies. In practice, any two products from the same industry or even the same company might be assigned to two different clusters. The point of clustering is to create enough different sets of equations to cover the whole spectrum of possibilities, so that, with pre-computed sets of structural equations in hand, a manager could easily use the decision tree to know which set of equations would likely be the best fit for the particular product under study.

4.3. Development of Structural Equations
As suggested in Wagner (1995), we used random data to vary model parameters and create a data set of experimental results (see Table 3). For our studies, we chose a setting with three marketing regions, six potential countries of operation (two in each region), 30 markets (19 in Europe, seven in Asia, and four in North America), and five time periods.

Low, medium, and high levels were chosen for four parameter sets: (i) fixed and investment (base) costs ($F_{mr}$, $E_{ins}$, and $A_{mn}$), (ii) amount of tariff, (iii) transportation cost per mile, and (iv) complexity costs ($B_{mn}$). A full-factorial experiment of these combinations amounts to $3^4 = 81$ trials (Table 4). We repeated this for 10 different random number seeds for a total of 810 trials per cluster to build the equations. Some of the solution times exceeded 3 hours, potentially leaving each cluster to vary model parameters and create a data set of experimental results (see Table 3). For our studies, we know which set of equations would likely be the best fit for the particular product under study.

Table 3 General Input Parameters Used in the Computational Studies*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base demand for country $j$</td>
<td>$DEM \times \left( \frac{\text{country } j's \text{ population}}{\text{US population}} \right) \times \left( \frac{\text{country } j's \text{ per capita income}}{\text{US per capita income}} \right)$</td>
</tr>
<tr>
<td>$D_{p,j}$</td>
<td>Uniform(0.75, 1.25) $\times$ (base demand for country $j$)</td>
</tr>
<tr>
<td>$\beta_n$</td>
<td>20% for all countries</td>
</tr>
<tr>
<td>$\tau_n$</td>
<td>Actual tax rate in each country</td>
</tr>
<tr>
<td>$\gamma_n$</td>
<td>34% in each country (to encourage no financing in the base scenarios)</td>
</tr>
<tr>
<td>Mileage</td>
<td>Actual mileage between countries</td>
</tr>
<tr>
<td>$V_{em}$</td>
<td>Uniform(0.8, 1.1) $\times$ VARSUB($h$)</td>
</tr>
<tr>
<td>$v_k$</td>
<td>Uniform(0.8, 1.1) $\times$ VARDC</td>
</tr>
<tr>
<td>$\rho_{mnk}$</td>
<td>$2 \times 1.1 \times$ VARSUB($h$)</td>
</tr>
<tr>
<td>$p_i$</td>
<td>$2 \times \left( \Sigma \rho_{mnk} + 1.1 \times \Sigma \text{VARDC} \right)$</td>
</tr>
<tr>
<td>Line and DC capacities</td>
<td>Maximum possible worldwide demand</td>
</tr>
<tr>
<td>$G_m$</td>
<td>1 square foot for each subassembly</td>
</tr>
<tr>
<td>Plant capacity</td>
<td>$\Sigma (G_m \times \text{maximum possible worldwide demand})$</td>
</tr>
<tr>
<td>$F_m$</td>
<td>Uniform(0.8, 1.2) $\times 0.01 \times DEM \times \Sigma [p_{mnk} - \text{VARSUB($h$)}]$</td>
</tr>
<tr>
<td>$l_m$</td>
<td>Uniform(0.8, 1.2) $\times 0.05 \times \Sigma [p_{mnk} - \text{VARSUB($h$)}]$</td>
</tr>
<tr>
<td>$u_k$</td>
<td>Uniform(0.8, 1.2) $\times 0.01 \times DEM \times \left( p_i - (\text{VARDC} + \Sigma p_{imk}) \right)$</td>
</tr>
<tr>
<td>$E_{sm}$</td>
<td>Uniform(0.01, 1) $\times 0.01 \times DEM \times \left( p_i - \text{VARSUB($h$)} \right)$</td>
</tr>
<tr>
<td>$e_{sm}$</td>
<td>Uniform(0.01, 1) $\times 0.05 \times \left[ p_{mnk} - \text{VARSUB($h$)} \right)$</td>
</tr>
<tr>
<td>$A_m$</td>
<td>Uniform(0.01, 1) $\times 0.01 \times DEM \times \left( p_i - \Sigma p_{mnk} - \text{VARSUB($h$)} \right)$</td>
</tr>
<tr>
<td>$a_m$</td>
<td>Uniform(0.01, 1) $\times 0.05 \times \Sigma [p_{mnk} - \text{VARSUB($h$)}]$</td>
</tr>
<tr>
<td>$h_k$</td>
<td>Uniform(0.01, 1) $\times 0.01 \times DEM \times \left( p_i - (\text{VARDC} + \Sigma p_{imk}) \right)$</td>
</tr>
<tr>
<td>$b_k$</td>
<td>Uniform(0.01, 1) $\times 0.05 \times \Sigma [p_{mnk} - \text{VARSUB($h$)}]$</td>
</tr>
<tr>
<td>$B_{sm}$</td>
<td>$\left(2 \text{total no. of subassemblies} \times \text{Uniform(0.8, 1.2)} \times 0.02 \times DEM \times \Sigma [p_{mnk} - \text{VARSUB($h$)}]\right)$</td>
</tr>
<tr>
<td>$\text{Tariff between countries from different regions}$</td>
<td>$0.05 \times p_i$</td>
</tr>
<tr>
<td>$\text{Transportation cost}$</td>
<td>Fixed transportation cost + transportation cost per mile $\times$ number of miles $\times$ tariff</td>
</tr>
<tr>
<td>$\text{Subassembly transportation cost fraction } \lambda_i$</td>
<td>VARSUB($h$)/(VARSUB($h$) + VARDC)</td>
</tr>
<tr>
<td>$\text{Depreciation } \delta_i$</td>
<td>Straight line</td>
</tr>
</tbody>
</table>

* VARSUB($h$) = average variable cost for component $i$; VARDC = average variable final assembly cost; and DEM = average US demand.

(except for three trials in cluster 8), producing very low adjusted $R^2$ structural equation values for SCATTER for those two clusters (0.065 and 0.181, respectively). Thus, for those two clusters, we set all SCATTER values to zero (and changed the structural equations accordingly).

The resulting equations (Appendix B) are designed to represent significant interaction effects and nonlinearities in a linear fashion. In that regard,
interpretations and comparison of the equations themselves may prove difficult, especially because the independent variables themselves contain multiple interactions. Nevertheless, we will conclude this section with a few sample observations.

Let us compare clusters 1 and 2, which represent the largest variable production cost differences in the top half of our decision tree. For GLOBAL, the coefficients for TRANS1 \(^{-1}\) are directly comparable. We observe that the same absolute level of increase for TRANS1 will provide a larger movement toward a regional strategy for the cheaper product (cluster 1). For FLEX, both clusters have terms for the two first-order COM variables (COM and 1/COM), and in each case the magnitude of the coefficient differences within the cluster suggests that the 1/COM term dominates. And for that term, cluster 1 has a much higher coefficient, suggesting that the same absolute increase in complexity cost should cause a more dramatic movement toward a dedicated strategy for the cheaper product. For SCATTER, the only first-order terms that are directly comparable are the COM-TRANS2 terms. For a given level of COM, we observe that the same absolute level of increase for TRANS2 will provide a larger movement toward a dedicated strategy for the cheaper product.

### 4.4. A Numerical Example

In this section, we present a numerical example that calculates the actual dependent variable values from the MIP model for a specific product and then takes it through the process of determining the appropriate cluster, determining the values of the independent variables, and calculating estimates of dependent variables. In section 5.2, we will illustrate how, over time, managers can use the structural equations to gauge the impact of certain shifts in economic conditions on the appropriateness of the network structure.

For this example, we chose a product (similar to a computer) with three subassemblies. We incorporated the following baseline parameters: variable subassembly costs = $150, $65, and $160, variable DC cost = $50, annual US demand = 100,000, and transnational transportation costs per unit = $8.50 plus $0.05 per mile. We used the same set of countries and markets as described in section 4.3. The input parameters from Table 3 were computed using, where applicable, the mean of the respective random distributions. Furthermore, all “Medium” parameter values from Table 4 applied.

We ran the full MIP model using the above parameters. One country in each region (Romania, Mexico, and Taiwan) each built two subassembly plants, the first producing subassemblies 1 and 3, and the second producing subassembly 2. With this network configuration of subassembly plants, the actual dependent variable values are as follows.

\[
GLOBAL = \frac{1}{5-1} \left( \frac{3}{5} \times \left( 1 - \min \left\{ \frac{618.350}{618.350}, \frac{471.680}{471.680} \right\} \right) + \frac{1}{5} \times \left( 1 - \min \left\{ \frac{558.900}{558.900}, \frac{471.680}{471.680} \right\} \right) \right) = 0.0000,
\]

\[
FLEX = \frac{(9/6)^{-1}}{3-1} = 0.2500,
\]

\[
SCATTER = \frac{3 \times \left[ \frac{1}{3-1} \max \left\{ 0, \left( \frac{1}{\max(1,1.1)+\min(1,1.1)+\min(1,1.1)} - 1 \right) \right\} \right]}{\min(1,1) + \min(1,1) + \min(1,1)} = 0.0000.
\]

Thus, the firm has a pure regional market focus, a plant focus that is mostly dedicated (the expression in parenthesis of FLEX reveals an average of 1.5 subassemblies per plant) but with some flexibility, and a pure concentrated network dispersion.

Table 7 presents the summary parameters that are used as inputs to the independent variables, which are then computed via Equations (1)–(4) as follows.
equalsto$$356.25/$8.50 = 41.9$$, which is $$> 20$$. Finally, as $$SUMV = $356.25 < $500$$, this product is classified in cluster 3.

Using the structural equations for cluster 3 (Appendix B), the estimated dependent variable values are:

$$GLOBAL = 1.521 - 15.461(0.101751)$$
$$+ 61.032(0.101751)² - 102.433(0.101751)³$$
$$+ 61.188(0.101751)⁴$$
$$- 0.521(0.101751)⁻¹(0.036866)$$
$$+ 1.367(0.101751)⁻¹(0.036866)² = 0.3079,$$

$$FLEX = -3.372 + 2.869(1.003272)⁻¹$$
$$+ 0.068(1.003272)⁻⁶ + 0.981(1.003272)$$
$$- 8.971(0.008954)(1.003272)⁻³$$
$$+ 105.624(0.008954)(1.003272) = 0.4675,$$

$$SCATTER = -0.01 + 0.033(0.008954)⁻¹(0.036866)$$
$$- 0.0000555(0.008954)⁻²(0.036866)$$
$$+ 0.000000286(0.008954)⁻³(0.036866)$$
$$- 0.01(0.008954)⁻¹(0.036866)² = 0.1003.$$
attempted to validate the model by examining the optimal profits from the MIP, where each dependent variable was restricted to being within a band (± 20% or ± 10%) around the structural equation estimates. These profits were compared to the unrestricted optimal profits from the MIP, as well as to a “lower bound.” The idea behind the lower bound was to illustrate what might happen if the manager chose the wrong facility network structure and then created the optimal location/allocations decisions within the framework of that structure. Specifically, to create the lower bound, we ran the MIP eight times, each time restricting the dependent variables of the network (i.e., the values of Equations (5)–(7) after running the MIP) to represent the eight “pure” combinations of our network classification system: 0-0-0, 0-0-1, 0-1-0, 0-1-1, 1-0-0, 1-0-1, 1-1-0, and 1-1-1. The lowest profit among the eight pure structures was designated as the lower bound. Clearly, other hybrid structures (e.g., 0.50, 0.75, 0.90) might produce even lower optimal profits; however, our goal in this exercise was for comparison purposes to see how the profits from creating a network structure near our structural equation estimates might compare to both the optimal solution and to a solution that was based on a poor strategic network choice. The results suggest that selecting the wrong facility network configuration was fairly well represented.

For these validation tests we created a brand new set of product cost structures, different from those used to create the structural equations (more details are available from the authors). In addition, to ensure that each dependent variable would cover the full 0–1 range, we created similar sorts of high/low experimental designs as used to create the equations, but we intentionally used different distributions (primarily NORMAL with different means and variances) than those used in the data-creation stage (Table 3). Also, the HIGH/LOW levels from Table 4 were all different for the validation set. The initial variable costs, along with some of the sales prices, transportation costs, demands, and fixed costs, were based on industry data from such sources as Isuppli Market Intelligence (2011), Ivey Case(a) (1996), Ivey Case(b) (2001), Ivey Case(c) (2008), and Wenzhou University (2010). These sources provided sufficient cost detail and a broad enough range of cost structures to populate our model and cover the eight clusters.

We ran at least 16 trials (different product cost structures) for each of our eight clusters, plus some extras using entirely different products for clusters 3 and 5 to bring the total up to 150. We ran 11 MIPs for each of these (optimal, ± 20%, ± 10%, and the eight pure strategies), representing 1650 total runs of the MIP. After running the optimal solutions for the 150 products, each of our dependent variables had a range from 0 to 1, so the full spectrum of possible network configurations was fairly well represented.

Table 8 displays the validation results. (Note that in a few categories the ± 10% bands outperformed the ± 20% bands. This occurred when we had to discard a few trials for the ± 10% bands because they were infeasible—i.e., trials where no configuration of plants could mathematically fall within a ± 10% band of the estimates.) Overall, the results indicate that using the structural equations generally provides a good estimate of the network structure, particularly when used as a monitoring tool. Network structures based on estimates provided from the equations produced close to optimal profits, especially when compared to poorly chosen network structures. In fact, over our full 150 experiments, nearly 95% of the profit loss that might have occurred by choosing the wrong pure network strategy is eliminated by basing the network strategy on a ± 20% range around our structural equation estimates.
5.2. Monitoring the Global Network Over Time—The Numerical Example Revisited

An important use of the dependent variable measures would be a process of recalculating them over time as conditions change. Here, we revisit the numerical example from section 4.4 to illustrate how managers can monitor their global network structures. Over time, managers can use the structural equations to gauge the impact of certain shifts in economic conditions on the appropriateness of the network structure.

Below, we provide examples that drive changes for each dependent variable. We reemphasize that estimated values are not meant to exactly match the actual values—but changes in estimated values should suggest a potential desire to investigate the economic viability of the current network structure and provide a signal as to the direction that a potential modified structure should take.

5.2.1. Market Focus. Case 1: Suppose that trade agreements are reached that eliminate all tariffs and, at the same time, either fuel prices drop or other transportation efficiencies cause the transportation cost per mile to be cut in half. The new estimated value for GLOBAL would rise to 0.7603, a large enough increase to suggest further investigation into the desirability to include more globalization in the market focus. Indeed, the optimal solution from the rerun MIP model recommends closing all dedicated plants in Mexico and transferring that demand to the Romanian plant. The new actual value for GLOBAL is 0.5729, and the optimal solution from the rerun MIP model again recommends closing all facilities in Mexico and transferring that demand to the Romanian plant. The new actual value for GLOBAL is again 0.5000.

5.2.2. Plant Focus. Case 1: Suppose that technology has changed and management has matured enough to the point where the firm expects to see no significant complexity costs in the future for producing its subassemblies at the same plant $(B_2 = 0)$. The new estimated value for FLEX would rise to 0.8295, a large enough increase to suggest further investigation into the desirability to include more flexibility in the plant focus. Indeed, the optimal solution from the rerun MIP model recommends closing all dedicated facilities in the three countries of production, leaving one flexible subassembly plant in each region. The new actual value for FLEX is 1.0000.

Case 2: Going in the other direction, suppose instead that complexity costs have doubled due to managing increasingly sophisticated technological requirements. The new estimated value for FLEX would fall to 0.2495, a large enough decrease to suggest further investigation into the desirability of more dedicated plants. Indeed, the optimal solution from the rerun MIP model recommends replacing all flexible plants in the three countries of production, resulting in a separate dedicated subassembly plant for each subassembly in each region. The new actual value for FLEX is 0.0000.

5.2.3. Network Dispersion. Case 1: Suppose that worldwide fuel prices skyrocket, resulting in a tripling of the per-mile transportation cost. The new estimated value for SCATTER would rise to 0.3118, a large enough increase to suggest further investigation into the desirability to include more scattered...
facilities (closer to the customers) in the network. Indeed, the optimal solution from the rerun MIP model recommends opening two new subassembly plants in the United States. The new actual value for SCATTER is 0.3333. If transportation costs jumped even more to five times the original cost per mile, the new estimated value for SCATTER would rise to 0.5112, and the new actual value would rise to 0.6667 because the model would recommend opening two new subassembly plants in Ireland (in addition to the new US plants).

Case 2: A demand increase could also help to justify a more scattered network dispersion strategy (building more plants closer to customers). Suppose that along with a tripling of the transportation cost per mile, demand for the product will escalate to a level of 10 times the current demand in each market. The new estimated value for SCATTER would rise to 0.6797, and the optimal solution from the rerun MIP model now recommends operating two subassembly plants in all six potential countries of production (now including India). The new actual value for SCATTER is 1.0000.

6. Conclusion

A profit-maximizing global network structure decision encompasses numerous issues requiring a significant amount of supporting data and a complicated mathematical programming approach. Even if a large mathematical program is solvable, its use in industry will be limited due to time and expertise requirements on management. In an attempt to make the power of mathematical modeling more accessible to managers, we have developed a predictive structural equations modeling approach to support management’s decision making by providing a guideline for the strategic structure of global networks based on our conceptual framework (Figure 1), whereby managers need only calculate a few relatively simple, yet explanatorily powerful, independent variables. Managers can capture the essence of the strategic network structure decision by using less detail than a traditional complete plant location model would necessitate. In an ongoing basis, as costs change, the equations can be used as a tracking device to quickly verify the appropriateness of the current network structure. This can allow for a decoupling of the strategic network decision and the actual plant location decision for a given network structure. Furthermore, either at the initial decision-making stage or in an ongoing monitoring basis, a manager might be able to use indications from the equations to add strategic-level constraints to a large MIP model to make it easier to solve.

We further describe a process using cluster analysis to create more robust sets of equations. If properly designed, such an approach should allow the manager of an applicable product to classify it into its appropriate cluster and then monitor potential changes to the strategic global network easily, without the hassles of generating a separate set of equations for that specific product.

As with any regression application, managers should be careful about potential limitations of predictive global sensitivity analysis. For example, the robustness of the equations may be limited by the input variations used to generate them, so care should be taken in estimating outputs particularly for inputs that lie beyond the boundaries of those used during equation generation. In particular, to use the specific equations developed in this article, products should fit within the modeling environment and parameter ranges described herein. Furthermore, by design, the independent variables may encompass summaries or approximations of several input parameters in the original model; hence, information will clearly be lost in the translation. Moreover, the terms that survive the stepwise regression elimination are completely data driven. Whereas the original independent variables themselves may have theoretical justification for being in the final equations, combinations such as $x^2y/z^2$ or $x^{-6}$ may have no particular theoretical justification and may be difficult to interpret. Finally, we view the primary use of structural equations as more of a monitoring tool for complicated environments rather than as a tool to make specific detailed decisions. In particular, if the equations imply that a change may be needed, at that point the manager may need to return to the detailed model (possibly constrained or directed by structural equation indications) to make every specific decision needed.

Numerical tests on our equations suggested that they generally provided a good estimate of the appropriate general network structure. In fact, the average deviation from optimal profit was only 1.25% in our experiments when we restricted each dependent variable to be within a ± 20% band around the structural equation estimates. We see this structural equations approach as an effective bridge between complicated mathematical modeling and real-world considerations. This approach could be extended to a wide range of applications, including (i) global network decisions for other supply chain structures, (ii) aggregate planning decisions, or perhaps (iii) certain resource allocation decisions.

Acknowledgments

The authors are very thankful to the senior editor and the referees whose very detailed reviews and suggestions helped to significantly improve this article. We also appreciate the volunteer efforts of Xiaotian Geng and Chieh.
Lee who provided some background research during the revision process. The research of Dr. Yang was supported by National Natural Science Foundation of China (No. 71001086).

Appendix A. MIP Model for Facility Location

A.1. Model Assumptions

- There are two facility levels: plants (producing subassemblies) and DCs (final assembly).
- The plants and DCs remain open throughout the finite planning horizon.
- The firm is a price taker in each market. All prices are quoted in local currency and then translated into common currency (say $) by using the real exchange rate.
- The market demands, selling prices, and transfer prices are independent of the structure of the facility network.
- At most, one DC (assumed to have adequate capacity to cover any one country’s demand) is allowed in each country. The rationale behind this assumption is that as shipment cost of a subassembly or final product within a country is assumed constant in our model, the opening of a second DC will result in additional fixed costs with no transportation cost savings. (Geographically large countries can easily be split into “subcountries,” as needed, to coincide with this assumption.)
- Similarly, any type of subassembly can be produced in at most one plant in any given country. Again, geographically large countries can easily be split into an arbitrarily large number of “subcountries,” as needed, to coincide with this assumption.
- Costs, prices, tax rates, interest rates, plant capacities, and discount rates remain constant over time. Demand and depreciation rates, however, may vary over time.
- Annual fixed costs and up-front investment costs are an increasing function of capacity and are modeled as a base cost plus a linear function of capacity.
- Capacities and costs are modeled for both plants and production lines within the plants. Production line variable investment and operating costs are linear functions of the number of subassemblies produced on those lines. Plant variable investment and operating costs are linear functions of the size of production lines (expressed as space requirements in square feet) installed in the plants.

A.2. Notation

<table>
<thead>
<tr>
<th>Indices</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$</td>
</tr>
<tr>
<td>$n$</td>
</tr>
<tr>
<td>$k$</td>
</tr>
<tr>
<td>$m$</td>
</tr>
<tr>
<td>$q$</td>
</tr>
<tr>
<td>$j$</td>
</tr>
<tr>
<td>$t$</td>
</tr>
<tr>
<td>$z$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{im}$</td>
</tr>
<tr>
<td>$v_k$</td>
</tr>
<tr>
<td>$F_m$</td>
</tr>
<tr>
<td>$f_m$</td>
</tr>
<tr>
<td>$B_{zm}$</td>
</tr>
<tr>
<td>$U_k$</td>
</tr>
<tr>
<td>$u_k$</td>
</tr>
<tr>
<td>$K_m$</td>
</tr>
<tr>
<td>$G_m$</td>
</tr>
<tr>
<td>$c_q$</td>
</tr>
<tr>
<td>$G_{sm}$</td>
</tr>
<tr>
<td>$E_{im}$</td>
</tr>
<tr>
<td>$e_{im}$</td>
</tr>
<tr>
<td>$A_m$</td>
</tr>
<tr>
<td>$a_m$</td>
</tr>
<tr>
<td>$H_k$</td>
</tr>
<tr>
<td>$h_k$</td>
</tr>
<tr>
<td>$S_{ij}$</td>
</tr>
<tr>
<td>$i, S_{mk}$</td>
</tr>
<tr>
<td>$D_p$</td>
</tr>
<tr>
<td>$r_n$</td>
</tr>
<tr>
<td>$R(t_n, T)$</td>
</tr>
<tr>
<td>$p(t_n, T)$</td>
</tr>
<tr>
<td>$τ_n$</td>
</tr>
<tr>
<td>$P_j$</td>
</tr>
</tbody>
</table>
Appendix A.2 Continued

\[ p_{mk} \text{ Transfer price of subassembly } i \text{ made in country } m \text{ shipped to DC in country } k \]
\[ d_{nt} \text{ Applicable depreciation rate in country } n \text{ in period } t \]
\[ \beta_n \text{ Discount rate of after tax cash flows in country } n \]
\[ l_n \text{ Maximum loan that country } n \text{ can give to the firm} \]

Decision variables

\[ x_{ijt} \text{ Units of subassembly } i \text{ made in country } m \text{ at plant } q \text{ and assembled in country } k \text{ during } t \]
\[ x_{kjt} \text{ Units assembled at a DC in country } k \text{ and sold in market } j \text{ during } t \]
\[ L_n \text{ The loan that the company will take from the country } n \text{ government for investment in } n \]
\[ Z_{imq} \text{ = 1 if a production line for subassembly } i \text{ is built in country } m \text{ at plant } q \text{ and 0 otherwise} \]
\[ Y_{imq} \text{ = 1 if a } q \text{th subassembly is produced in country } m \text{ at plant } q \text{ and 0 otherwise} \]
\[ y_k \text{ = 1 if a DC is operated in country } k \text{ and 0 otherwise} \]
\[ W_{imq} \text{ Size (in units) of production line } i \text{ built in country } m \text{ at plant } q \]
\[ Q_{mq} \text{ Size (in square feet) of plant } q \text{ in country } m \]
\[ W_k \text{ Size (in units) of a DC operated in country } k \]

A.3. The Objective Function

Revenue from units assembled and subassemblies produced in country \( n \) in period \( t \):

\[ R_{nt} = \sum_{j=1}^{J} p_{ij} x_{ijt} + \sum_{i=1}^{I} \sum_{q=1}^{Q} \sum_{k=1}^{K} p_{i} x_{kjt} \cdot \]

Variable production costs and cost of goods sold in country \( n \) in period \( t \):

\[ \begin{align*}
\delta_{nt} &= \sum_{j=1}^{J} \nu_{ij} x_{ijt} + \sum_{i=1}^{I} \sum_{q=1}^{Q} \sum_{k=1}^{K} p_{i} x_{kjt} \\
&\quad + \sum_{i=1}^{I} \sum_{q=1}^{Q} \sum_{k=1}^{K} V_{in} x_{kjt}. 
\end{align*} \]

Annual fixed costs of a DC and subassembly plants in country \( n \):

\[ \gamma_n = U_n y_n + u_n w_n \]
\[ + \sum_{q=1}^{Q} \left( f_n y_{1mq} + f_n q_{nq} + \sum_{z=2}^{I} B_{zn} y_{znq} \right). \]

Transportation costs from facilities (plants and DCs) located in country \( n \) in period \( t \):

\[ \eta_{nt} = \sum_{j=1}^{J} S_{nj} x_{njt} + \sum_{i=1}^{I} \sum_{q=1}^{Q} \sum_{k=1}^{K} \lambda_{ik} S_{nk} x_{kjt}. \]

Loan payment in country \( n \) in period \( t \):

\[ \phi_{nt} = L_n R(t_{nt}, T). \]

Loan interest payment in country \( n \) in period \( t \):

\[ \phi^i_{nt} = L_n \rho t (t_{nt}, T). \]

Investment cost in country \( n \):

\[ \xi_n = H_n y_n + h_n w_n + \sum_{q=1}^{Q} (a_n y_{1mq} + a_n q_{nq}) \]
\[ + \sum_{i=1}^{I} \sum_{q=1}^{Q} (E_{in} Z_{imq} + e_{in} W_{imq}). \]

Depreciation expense in country \( n \) in period \( t \):

\[ \psi_{nt} = \xi_n d_{nt}. \]

Before-tax income in country \( n \) in period \( t \):

\[ \pi_{nt} = \xi_n - (\delta_{nt} + \gamma_n + \eta_{nt} + \phi^i_{nt} + \psi_{nt}). \]

Corporate income tax paid in country \( n \) in period \( t \):

\[ \omega_n = \pi_{nt} r_n. \]

Cash expenditures in fixed assets in year \( 0 \) that are not financed by external sources:

\[ \mu = \sum_{n=1}^{N} (\xi_n - L_n). \]

Adding back loan interest and depreciation expenses to cash flows, the objective function which maximizes the net present value is:

\[ \text{OBJ} = \max \left[ \sum_{n=1}^{N} \sum_{t=1}^{T} \left( \pi_{nt} + \phi^i_{nt} + \psi_{nt} - \phi_{nt} - \omega_n \right) \left( 1 + \beta_n \right)^{-t} \right] - \mu. \]

A.4. The Set of Constraints

The constraints are:

1. Demand constraints: \( \sum_{k=1}^{K} X_{imqt} \leq D_{jt} \quad \forall j, t \)

2. Subassembly line capacity:

\[ \sum_{k=1}^{K} X_{imqt} \leq W_{imq} \quad \forall i, m, q, t, \]
\[ W_{imq} \leq K_{im} Z_{imq} \quad \forall i, m, q, \]
\[ \sum_{q=1}^{Q} Z_{imq} \leq 1 \quad \forall i, m, \]
\[ \sum_{i=1}^{I} z_{im(q+1)} \leq \sum_{i=1}^{I} Z_{imq} \quad \forall m, (q < I) \]

(for solution efficiency),

\[ Z_{imq} \mid (q > i) = 0 \quad \forall i, m \] (for solution efficiency).
3 Plant capacity constraints:
\[
\sum_{i=1}^{I} G_{im} W_{imq} \leq Q_{mq} \quad \forall m, q,
\]
\[
Q_{mq} \leq C_{m} Y_{1mq} \quad \forall m, q,
\]
\[
Y_{1m(q+1)} \leq Y_{1mq} \quad \forall m, (q < I) \quad \text{(for solution efficiency)}.
\]

4 DC capacity constraints:
\[
\sum_{j=1}^{J} x_{kjt} \leq w_{k} \quad \forall k, t,
\]
\[
w_{k} \leq c_{k} y_{k} \quad \forall k.
\]

5 Conservation of subassembly flows
\[
\sum_{m=1}^{M} \sum_{n \neq 1}^{N} X_{imnkt} = \sum_{j=1}^{J} x_{kjt} \quad \forall i, k, t.
\]

6 Loan ceilings: \( L_{n} \leq I_{n} \quad \forall n \) and \( L_{n} \leq \zeta_{n} \quad \forall n \).

7 Non-negative profit in each country in each period (assumption of convenience to avoid unnecessarily complicated tax calculations):
\[
\pi_{nt} \geq 0 \quad \forall n, t.
\]

8 Count the number of different subassemblies produced at each plant:
\[
\sum_{m=1}^{M} Z_{imq} = \sum_{z=1}^{Z} Y_{zmq} \quad \forall m, q,
\]
\[
Y_{(z+1)mq} \leq Y_{zmq} \quad \forall z \neq I, m, q.
\]

9 Specification of Decision Variables:
\[
X_{imnkt}, x_{kjt}, L_{n}, W_{imq}, Q_{mq}, w_{k} \geq 0 \text{ for all indices},
\]
\[
Z_{imq}, Y_{zmq}, y_{k} = \text{ binary for all indices}.
\]

Notes: The plant index \( q \) is necessary because, under the model assumptions, a country could have between 1 and \( I \) total subassembly plants, depending upon the degree of plant focus. Because numerous sets of equivalent variables could be created simply via different indexing sequences, we eliminated redundant combinations by incorporating three additional sets of constraints to essentially disallow searches along completely redundant paths. First, the fourth set of constraints under “Subassembly line capacity” simply ensures that the plants that are used are numbered consecutively starting with 1. For example, for a country producing five subassemblies in a total of three plants, the plants would be numbered 1, 2, and 3, thus eliminating the use of indices 4 and 5 for \( q \) for that country. The same logic applies to the third set of constraints under “Plant capacity constraints.” Finally, the fifth set of constraints under “Subassembly line capacity” reduces the possible combinations of plant index numbers used by forcing the value of a plant index to be less than or equal to the index values of all subassemblies produced in that plant. For example, consider a country with three subassemblies, where subassemblies 2 and 3 are produced at the same facility. According to these constraints, subassembly 1 must be produced at plant 1, and subassemblies 2 and 3 must be produced at plant 2. The program will not be allowed to search along the redundant path of reversing those plant index designations.

The purpose of constraint set 7 (“Non-negative profit in each country in each period”) is simply to prevent the model from playing “tax games” that might imply, for example, that a government should pay the firm for losing money during a period. Sometimes a firm can alter transfer prices paid to itself to take advantage of tax differences among countries. Such considerations invoke complicated legal and accounting issues that are beyond the scope of this research. Given that our model fixed the transfer prices ahead of time, the primary effect of such a set of constraints would be to not satisfy potential demand in certain countries in case such satisfaction would be cost prohibitive. For all of the experiments that we ran, we never came across a case where the company should not produce anything at all.

Appendix B. Structural Equations for the Eight Clusters

Note: the \( p \)-value for each coefficient is provided in order after each equation.

Cluster 1:
\[
GLOBAL = -0.089 + 0.033TRANS1^{-1} - 0.018TRANS1^{-2} \cdot TRANS2
- 0.0000536TRANS1^{-3} \cdot TRANS2 + 0.0000000768TRANS1^{-1} \cdot TRANS2^{-2}
- 0.000000000373TRANS1^{-1} \cdot TRANS2^{-3}[0, 0, 0, 0, 0, 0]
\]
\[ FLEX = -12.537 + 9.875COM^{-1} + 0.003EOS^{-1} \cdot COM^{-2} + 0.428COM^{-6} + 3.752COM - 0.005COM^6 \\
- 1.055COM^{-5} - 0.004EOS^{-1} \cdot COM^{-3} + 8.036EOS \cdot COM^3 - 5.751EOS \cdot COM^{-1} \\
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0] \]

\[ SCATTER = -0.138 + 10.17COM \cdot TRANS2 + 0.002EOS^{-1} \cdot TRANS2 - 7.819COM^2 \cdot TRANS2 \\
- 2106.061EOS^3 \cdot TRANS2 + 1.798COM^3 \cdot TRANS2 - 1165.671TRANS2^6 [0, 0, 0, 0, 0, 0, 0] \]

Cluster 2:

\[ GLOBAL = 0.742 - 5.871TRANS1 + 13.828TRANS1^2 - 9.507TRANS1^3 \\
- 0.087TRANS1^{-1} \cdot TRANS2 + 0.02TRANS1^{-1} - 0.02TRANS1^{-2} \cdot TRANS2 [0, 0, 0, 0, 0.004, 0, 0] \]

\[ FLEX = -3.445 + 2.865COM^{-1} + 0.087COM^{-6} + 1.041COM - 7.2EOS \cdot COM^{-3} \\
+ 1113.823EOS^3 \cdot COM [0, 0, 0, 0, 0, 0] \]

\[ SCATTER = -0.067 + 6.279COM \cdot TRANS2 + 0.388EOS \cdot COM^2 - 3.235COM \cdot TRANS2^3 \\
+ 1.103COM^3 \cdot TRANS2 - 1519.801EOS^3 \cdot TRANS2 - 4.48COM^2 \cdot TRANS2 + 9.74E-20EOS^{-6} \\
[0, 0, 0.251, 0, 0, 0, 0, 0] \]

Cluster 3:

\[ GLOBAL = 1.521 - 15.461TRANS1 + 61.032TRANS1^2 - 102.433TRANS1^3 + 61.188TRANS1^4 \\
- 0.521TRANS1^{-1} \cdot TRANS2 + 1.367TRANS1^{-1} \cdot TRANS2 [0, 0, 0, 0, 0, 0, 0, 0] \]

\[ FLEX = -3.372 + 2.869COM^{-1} + 0.068COM^{-6} + 0.981COM - 8.971EOS \cdot COM^{-3} \\
+ 105.624EOS^2 \cdot COM^{-1} [0, 0, 0, 0, 0, 0, 0] \]

\[ SCATTER = -0.01 + 0.033EOS^{-1} \cdot TRANS2 - 0.0000555EOS^{-2} \cdot TRANS2 \\
+ 0.0000000286EOS^{-3} \cdot TRANS2 - 0.01EOS^{-1} \cdot TRANS2^2 [0.106, 0, 0, 0, 0] \]

Cluster 4:

\[ GLOBAL = 1.523 - 15.142TRANS1 + 56.839TRANS1^2 - 90.598TRANS1^3 \\
- 0.539TRANS1^{-1} \cdot TRANS2 + 49.742TRANS1^4 + 1.348TRANS2 [0, 0, 0, 0, 0, 0, 0, 0] \]

\[ FLEX = 0.792 + 4.31COM^{-2} + 0.003EOS^{-1} \cdot COM^{-1} + 0.384COM^{-6} - 0.001EOS^{-1} \\
- 1.046COM^{-5} - 0.003EOS^{-1} \cdot COM^{-2} - 3.93COM^{-1} [0, 0, 0, 0, 0, 0, 0, 0] \]

\[ SCATTER = -0.031 + 0.002EOS^{-1} \cdot TRANS2 + 3.066TRANS2 - 19.927EOS \cdot TRANS2 \\
- 0.004EOS^{-1} \cdot TRANS2^2 - 3.476TRANS2^3 [0, 0, 0, 0, 0, 0] \]

Cluster 5:

\[ GLOBAL = -0.017 + 0.096COM^{-1} \cdot TRANS2 + 179.248EOS^3 \cdot COM - 0.001COM^{-1} \cdot TRANS2^3 \\
+ 0.003COM \cdot TRANS2^2 + 5.247EOS^2 \cdot TRANS2 [0.061, 0, 0, 0, 0, 0] \]

\[ FLEX = 0.103 - 0.615COM^{-2} - 0.000000632EOS^{-2} \cdot COM^{-1} - 0.624COM^{-6} \\
+ 0.000000000294EOS^{-3} - 21.413EOS \cdot COM^{-3} + 246.534EOS^2 \cdot COM^{-1} + 1.684COM^{-5} \\
[0.081, 0, 0, 0, 0, 0, 0, 0] \]
$SCATTER = -0.084 + 0.000000000419EOS^{-3}.TRANS2^{-1} - 0.000000054EOS^{-1}.TRANS2^{-3}$
$+ 0.001EOS^{-1}.TRANS2^{-1} + 0.097COM.TRANS2^{-1} - 0.014COM.TRANS2^{-2}$
$+ 0.009COM^{-2}.TRANS1 - 0.000000943EOS^{-2}.TRANS2^{-1} + 0.001EOS.TRANS2^{-3}$
$- 0.00000411TRANS2^{-4}[0,0,0,0,0,0,0,0]$

Cluster 6:

$GLOBAL = 0.948 - 4.293TRANS1 + 3.958TRANS1^2 + 0.00000268TRANS1^{-1}[0,0,0,0.01,0.038]$ $FLEX = -0.392 + 0.835COM^{-2} - 0.001EOS^{-1}.COM^{-2} - 0.038COM^{-6} - 1.261EOS[0,0,0,0,0]$

$SCATTER = 0$

Cluster 7:

$GLOBAL = -0.089 + 0.033TRANS1^{-1} - 0.018TRANS1^{-1}.TRANS2$ $- 0.0000536TRANS1^{-3}.TRANS2 + 0.000000768TRANS1^{-1}.TRANS2^{-2}$ $- 0.0000000000373TRANS1^{-1}.TRANS2^{-3}[0,0,0,0,0,0]$

$FLEX = -1.253 + 1.453COM^{-2} - 0.00000962EOS^{-2}.COM^{-1} + 12.717EOS.COM^{-3} + 1.59E-19EOS^{-6}$ $+ 1471918.934EOS^6 - 2127.506EOS^3.COM^{-1} + 0.001EOS^{-1}.COM[0,0,0,0,0,0,0]$

$SCATTER = -0.013 - 4.538EOS.COM^2 + 2.94COM.TRANS^3 + 4.28COM.TRANS2$ $- 7.2588COM.TRANS^2 + 0.017EOS.TRANS^2 - 0.016COM^3 - 0.00003EOS.TRANS^2$ $+ 3.359EOS^2.TRANS^2 - 6.698EOS^2.TRANS[0,0,0,0,0,0,0]$

Cluster 8:

$GLOBAL = 0.978 - 3.887TRANS1 + 392.242TRANS1^6[0,0,0,0.021]$ $FLEX = 0.655 + 1.338COM^{-3} - 0.151COM^{-6} - 1.555COM{-1} - 1.65EOS.COM[0,0,0,0,0]$

$SCATTER = 0$

References


