MSOM MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 10, No. 1, Winter 2008, pp. 19–23
ISSN 1523-4614 | EISSN 1526-5498 | 08 | 1001 | 0019

informs®

DOI 10.1287/msom.1070.0176

Data Set

Real-World Multiechelon Supply Chains Used for Inventory Optimization

Sean P. Willems
School of Management, Boston University, Boston, Massachusetts 02215,
willems@bu.edu

This data set describes 38 multiechelon supply chains that have been implemented in practice. These chains exhibit special structure that can be used to inform and test analytical models. Although the data were not collected with the intention of econometric analysis, they may be useful in an empirical study. The data described in this paper are publicly available at the journal's website (http://msom.pubs.informs.org/ecompanion.html).

Key words: multiechelon inventory system; multistage supply chain application; inventory optimization; supply chain data set
History: Received: February 28, 2006; accepted: February 8, 2007. Published online in *Articles in Advance* December 11, 2007.

1. Introduction

The data set in the accompanying online appendix[1] sheds light on how practitioners equipped with an inventory optimization tool are modeling supply chains in reality. The paper's primary objective is to inform researchers developing analytical supply chain models by presenting supply chains that business users and consultants have created. Each chain contains the exact supply chain network created by the contributor, plus the exact characterization of leadtimes at every stage in the supply chain, plus realistic yet disguised values for all the supply chain's data. A secondary use of this data might be to support empirical research into the nature of the supply chains created. The data are open to all researchers as long as the researcher is willing to cite this document as the source. Section 2 presents background information for the chains and §3 presents some observations to place the data in context.

An online appendix accompanies this paper to provide a more detailed explanation of the data associated with each chain.

[1] The online appendix for this paper is available on the *Manufacturing & Service Operations Management* website (http://msom.pubs.informs.org/ecompanion.html).

2. High-Level Presentation of the Chains

The chains described in this paper comprise actual supply chain models created by either company analysts or consultants. Because models using the data have been implemented in practice, these chains demonstrate how real users have modeled their real-world supply chain situations.

2.1. Chain Selection Criteria

PowerChain Inventory (PCI) is a software application that optimizes inventory levels and locations across the supply chain. PCI is based on the guaranteed service model of inventory placement as described in Graves and Willems (2003). PCI's minimum data requirements include the bill of materials, the cost and time at every stage, and the demand characterization at demand stages (consisting of average demand per period, standard deviation per period, service level, and maximum time the customer is willing to wait). Given this data, PCI optimizes inventory cost across the network subject to the service requirements at demand stages.

Beginning in 2004, I asked companies using PCI to send me what they considered to be the most representative acyclic supply chain model they had created. Because several of the companies have many users who have created thousands of chains, the

Table 1 General Classification of the Data Set

Chain name	Company identifier	SIC code	SIC description
1	1	2869	Industrial organic chemicals, not elsewhere classified
2	2	3674	Semiconductors and related devices
3	3	3577	Computer peripheral equipment, not elsewhere classified
4	4	3944	Games, toys, and children's vehicles, except dolls and bicycles
5	5	2099	Food preparations, not elsewhere classified
6	6	3421	Cutlery
7	7	3531	Construction machinery and equipment
8	8	3845	Electromedical and electrotherapeutic apparatus
9	9	2043	Cereal breakfast foods
10	10	5064	Electrical appliances, television and radio sets
11	7	3531	Construction machinery and equipment
12	9	2043	Cereal breakfast foods
13	3	3674	Semiconductors and related devices
14	11	4731	Arrangement of transportation of freight and cargo
15	12	2841	Soap and other detergents, except specialty cleaners
16	8	3845	Electromedical and electrotherapeutic apparatus
17	13	3577	Computer peripheral equipment, not elsewhere classified
18	14	3577	Computer peripheral equipment, not elsewhere classified
19	13	3577	Computer peripheral equipment, not elsewhere classified
20	15	3577	Computer peripheral equipment, not elsewhere classified
21	16	2844	Perfumes, cosmetics, and other toilet preparations
22	17	2834	Pharmaceutical preparations
23	18	2851	Paints, varnishes, lacquers, enamels, and allied products
24	19	3546	Power-driven handtools
25	20	3523	Farm machinery and equipment
26	21	3724	Aircraft engines and engine parts
27	22	3845	Electromedical and electrotherapeutic apparatus
28	23	3572	Computer storage devices
29	24	3692	Primary batteries, dry and wet
30	25	4731	Arrangement of transportation of freight and cargo
31	20	3523	Farm machinery and equipment
32	16	2844	Perfumes, cosmetics, and other toilet preparations
33	12	2844	Perfumes, cosmetics, and other toilet preparations
34	26	3661	Telephone and telegraph apparatus
35	22	3845	Electromedical and electrotherapeutic apparatus
36	27	3523	Farm machinery and equipment
37	28	2869	Industrial organic chemicals, not elsewhere classified
38	29	3724	Aircraft engines and engine parts

requirements were kept simple to achieve my goal of generating a test bed of general acyclic supply chain networks that can support the development of analytical models. The two major requirements for each contributor were that they only submit a single chain and that the single chain represent the complexity that they have to model in their business. Table 1 summarizes high level details of the data set. In the nine cases where two supply chains come from the same company, the chains come from users in different business units.

It is fair to note that other criteria could have been suggested when requesting the chains. For example, one could have asked for the chain associated with the most successful supply chain improvement initiative at the company where success might be defined in terms of dollars saved, or dollar-volume of business affected, or number of employees influenced. Another valid selection criterion could have been the most illustrative chain in an effort to demonstrate the modeling aptitude present at the company. While these are valid and potentially interesting criteria, the goal in identifying a chain with representative complexity is to capture what business users consider to be chains worth modeling and optimizing.

2.2. Introducing the Chains

From a graph-theoretic perspective, a supply chain map forms a network of stages and arcs. A stage represents the processing activity of a stock keeping unit (SKU) at a location. For example, if two SKUs flow through a distribution center, the network will contain one stage for each SKU at the distribution center. A stage is a candidate location to hold safety stock; stages with no outgoing arcs are demand stages. Arcs denote the precedence relationship between stages. The chains in the data set vary from 8 to 2,025 stages and 10 to 16,225 arcs.

Because the submitted chains contain stage names that use company-specific acronyms or information-system-assigned labels, the first data synthesis step entails assigning each stage one of five classification labels that reflect the function the stage performed. Within each classification, stages are numbered sequentially starting at 0001. The stage classifications are noted in Table 2.

Table 3 specifies the data provided at each stage in the supply chain.

The online appendix to this paper defines each data element in detail. To highlight some summary statistics of the chains, the average cost of goods sold per unit varies from $3.12 to $150,816.00. The average supply chain length varies from 16.4 days to 524.3 days. Twenty-six of the 38 chains employ variable stage times, with 14 chains specifying some stage times as normally distributed, six chains specifying discrete random variables, and six chains using both characterization methods.

3. Observations About the Data

First, many chains exhibit special structure. As defined in Humair and Willems (2006), chains with clusters of commonality (CoC) can be transformed into a modified network that is itself a spanning tree. CoC chains reflect a network where there is a lot of commonality between a limited number of echelons but the rest of the chain is relatively disjoint. Half the submitted chains are CoC chains. In the chains that are not CoC chains, very few have parts that skip echelons. So, even among all the complexities of general acyclic networks, real-world supply chains exhibit special structure and look more like Chain 12 in Figure 1 than Chain 19 in Figure 2.

Table 2 Classifications Used to Label Every Stage in the Chains

Classifications label	Activity
Dist_	A stage that distributes an item
Manuf_	A stage that manufactures or assembles an item
Part_	A stage that procures an item
Retail_	A stage that acts as a demand orgination point
Trans_	A stage that transports an item between stages

Table 3 Stage Data

Data field	Applicable stages	Description
stageCost	All stages	Direct cost added at the stage
stageTime	All stages	Average processing time at the stage
stDevStageTime	All stages	Standard deviation of processing time at the stage
avgDemand	Only demand stages	Daily demand rate
stdDevDemand	Only demand stages	Daily standard deviation of demand
serviceLevel	Only demand stages	Percentage of orders stage plans to satisfy
maxServiceTime	Only demand stages	Maximum time the customer is willing to wait
relDepth	All stages	Relative depth of node to network root
stageClassification	All stages	Classification stage belongs to
xPosition	All stages	Horizontal distance from (0, 0) picture position
yPosition	All stages	Vertical distance from (0, 0) picture position

Second, most chains do not progress linearly through the classified processes of parts to manufacturing to transportation to distribution to retail. Figure 3 displays a subgraph from Chain 33 that captures an implementation of delayed differentiation. Near the top of Figure 3, Manuf_0078 is a packaging operation that puts together two different intermediate products. These products (of which Manuf_0072 is visible in the figure) are themselves distributed through a two-tier network and sold separately.

Third, chains of 150 and 350 stages demonstrate the greatest complexity. The intuition is that small chains are trivial and large chains replicate the bill of materials in an enterprise planning system. Users craft midsize chains to solve thorny supply chain problems. The average number of routings for chains of 50 and 150 stages is 5.38. Chains with more than 1,000 stages have on average 2 routings; chains with fewer than 100 stages have 1.83 routings on average. If one calculates the weighted average relative depth for a

Figure 1 **Chain 12 Is a CoC Network with Significant Commonality in Two Echelons**

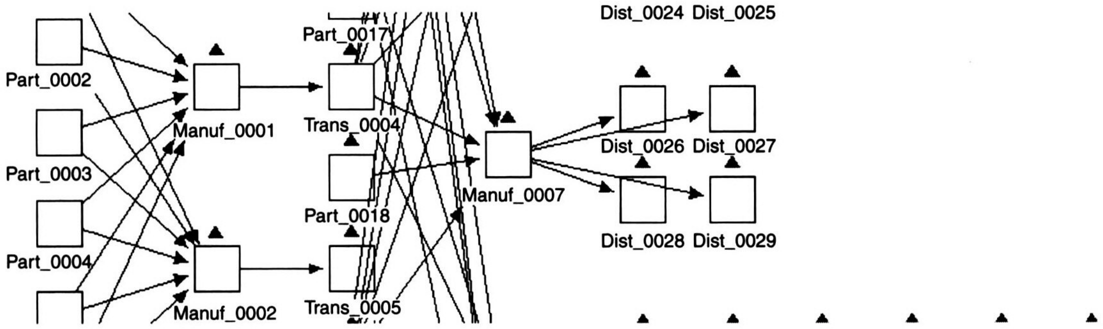


Figure 2 **Chain 19 Has Stages that Supply Stages in Multiple Echelons**

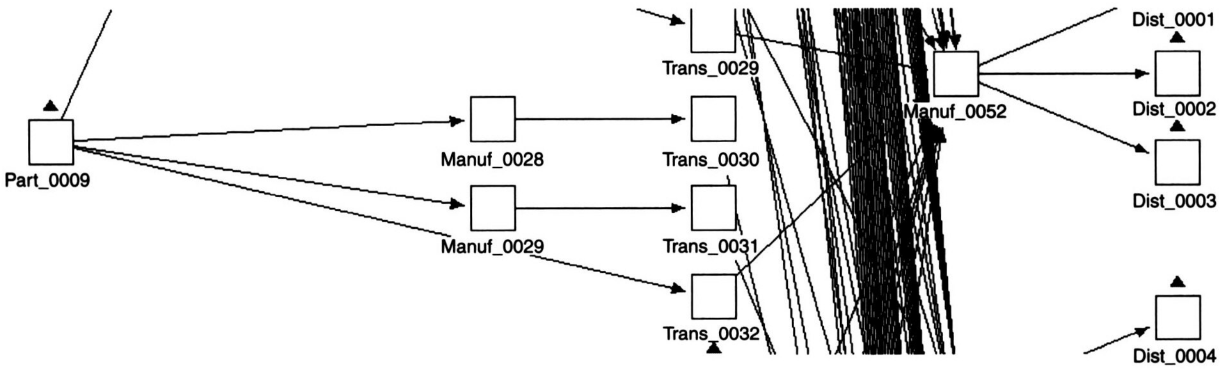


Figure 3 **Subgraph from Chain 33 Showing a Postponement Operation**

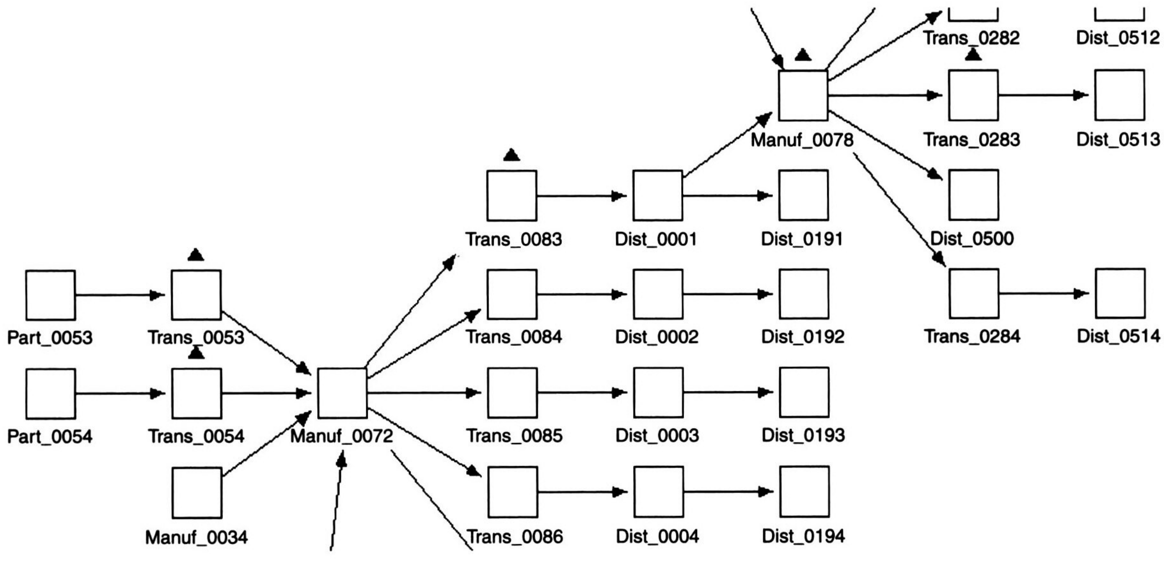
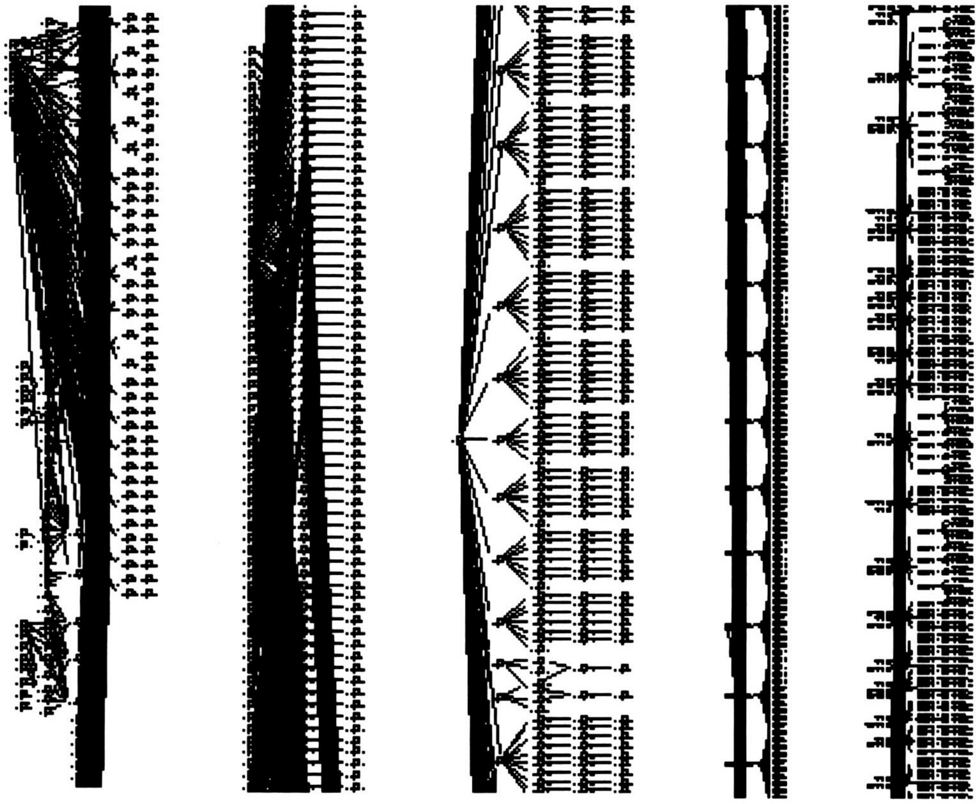


Figure 4 From Left to Right, Chains 24, 25, 30, 31, and 33 Are Shown



chain as the number of stages with each depth times that depth divided by the total number of stages, the average across chains with 150 to 350 stages is 2.56, which is again the largest value among the chains.

Fourth, within a chain, there are many redundant patterns. Even intricate chains such as Chain 24 in Figure 4 show a great deal of complexity followed by repeated patterns. Larger networks like Chains 25, 30, 31, and 33 are less complicated but still show dominant replicating patterns.

Finally, chains span multiple divisions within companies. Because the contributors submitted acyclic networks, it is not surprising that the chains are multiechelon. It is however interesting that there are very few manufacturing-centric chains such as Chain 10 or distribution-focused chains such as Chain 30.

Conclusions

Real-world supply chains exhibit special structure. To inform research, this data set presents 38 actual supply chains that have been used in practice. Each chain contains the exact supply chain network created by the contributor plus the exact characterization of lead times at every stage in the supply chain. While the data has been disguised, it has been kept consistent with the overall metrics of the contributed chain. This data set is designed to enable researchers to use a test bed of supply chain models to test the efficiency and effectiveness of their algorithms.

Acknowledgments

The author thanks the editor and three referees for their very helpful and constructive feedback on two earlier versions of the paper.

References

Graves, S. C., S. P. Willems. 2003. Supply chain design: Safety stock placement and supply chain configuration. A. G. de Kok, S. C. Graves, eds. *Handbooks in Operations Research and Management Science*, Vol. 11. *Supply Chain Management: Design, Coordination and Operation*, Chapter 3. North-Holland, Amsterdam, The Netherlands, 95–132.

Humair, S., S. P. Willems. 2006. Optimizing strategic safety stock placement in supply chains with clusters of commonality. *Oper. Res.* **54**(4) 725–742.