

Lawrence Berkeley National Laboratory

Recent Work

Title

A DYNAMIC PROGRAMMING APPROACH TO ESTIMATING GAPS IN LARGE ECONOMIC DATA BASES

Permalink

<https://escholarship.org/uc/item/1z80n37g>

Author

Andres, F.M.

Publication Date

1975-09-01

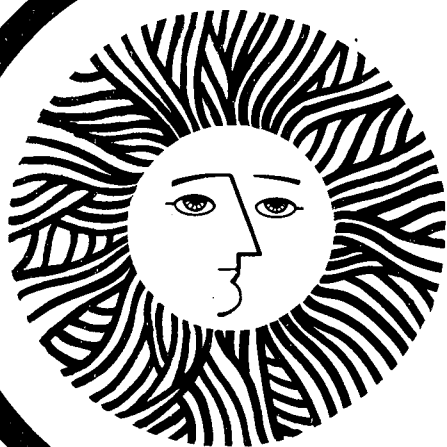
LBL-4271

c.1

RECEIVED
LAWRENCE
BERKELEY LABORATORY

MAR 30 1976

LIBRARY AND
DOCUMENTS SECTION



A DYNAMIC PROGRAMMING APPROACH TO ESTIMATING GAPS IN LARGE ECONOMIC DATA BASES

Energy and Environment Division
Lawrence Berkeley Laboratory
University of California

For Reference

Not to be taken from this room

LBL-4271

c.1

8 9 7 2 0 7 0 0 0 0

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Presented at Operations Research
Society of America Conference,
Las Vegas, Nevada, November 16-19, 1975

LBL-4271

A DYNAMIC PROGRAMMING APPROACH TO ESTIMATING
GAPS IN LARGE ECONOMIC DATA BASES

F. M. Andres, H. Ruderman and J. Sathaye

September 1975

Prepared for the U. S. Energy Research and
Development Administration under Contract W-7405-ENG-48

6 9 4 2 0 4 1 0 0 0 0

A DYNAMIC PROGRAMMING APPROACH TO ESTIMATING
GAPS IN LARGE ECONOMIC DATA BASES

F.M. Andres, H. Ruderman and J. Sathaye

Energy and Environment Division
Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

September 1975

1) MOTIVATION

Many enterprising modelers have developed precise mathematical models of social systems expecting equally precise input data from the economic census. They soon discover that, unfortunately, as the precision of the data increases so does the amount of missing data. First, and most commonly, known data is "withheld to avoid disclosing figures for individual companies."⁶ The number of withheld items becomes significant when the economy is divided into about three hundred or more industry groups for regions the size of states or smaller. This degree of disaggregation is often necessary to capture important interactions. Secondly, a major reason for missing data is that the same size for some categories is insufficient for reasonable confidence in the data. This occurs frequently in the Census of Transportation.⁷

However, there is usually enough other information about the missing data to constrain it and to estimate a probability distribution for it within the constraints. For example, the two data items showing the total value of shipments for the flat glass industry (SIC 321) and the pottery and related

products industry (SIC 326) are missing for California in the 1967 Census of Manufactures (see Table I). The former employs between 1,000 and 2,499 employees in 10 establishments and the latter employs more than 2,500 employees in 158 establishments. National tables (see Table II) provide data on the average character of each industry such as total establishments, value of shipments, and employees for the entire nation. From this information, a probability distribution of the missing data can be estimated. The sum of the two data items, SIC 321 and SIC 326, is found by subtracting the value of shipments for SIC numbers 322, 323, 324, 325, 327, 328 and 329 from the value of shipments for SIC 32. Although the value of shipments for all other states for SIC numbers 321 and 326 is now known^{*}, the sum of the California value and the other states' value for each is supplied in the national table. These two sums constrain the missing data items.

1.1) Problem Statement

It is proposed to find a maximum likelihood estimator of the missing data items. Assuming statistical independence of the data items, the probability of any particular set of estimates of missing items such as mentioned above becomes the product of the individual probabilities and taking logarithms yields a sum of functions. It is desired to maximize this sum of functions with the reported sum over all industries at the next aggregated SIC level in each region, and the reported national sum over regions for each industry being the constraints. Mathematically, the problem becomes,

* Although the value for all other states could hypothetically be found by adding values from other states, many will be missing. Thus, only a lower bound would be found. It is ignored herein for simplicity.

$$\max \sum_i \sum_j r_{ij}(x_{ij}) \quad (1)$$

$$\text{s.t. } \sum_j x_{ij} = a_i$$

$$\sum_i x_{ij} = b_j,$$

where i = Index of industry (SIC)

j = Index of region

x_{ij} = Missing value of the data item in region j , SIC i ,

$r_{ij}(x_{ij})$ = The logarithm of the probability density in region j , SIC i , when the missing value is x_{ij} . This will be called the reward function.

a_i = The reported national total for industry i (row total)

b_j = The reported regional total for the next aggregated SIC level (column total).

The problem is more readily visualized as one of finding maximal allocations to the elements of a matrix subject to specified row and column sums. With this visualization, the problem can be recognized as the classical transportation problem² with a nonlinear objective.

2) GENERAL DESCRIPTION OF THE ALGORITHM

To solve the problem stated, a suboptimal dynamic programming algorithm which is an extension of the classical "allocation problem"¹ is presented. The terminology of Bellman and Dreyfus¹ will be used where applicable. The stages of the problem are "L's" of the matrix where the i^{th} "L" = $L_i = \{x_{1i}, x_{2i}, \dots, x_{ii}, x_{i(i-1)}, x_{i(i-2)}, \dots, x_{i1}\}$ (see Figure 1). Assuming the matrix is $m \times n$ then the number of stages is $k_1 = \min(m, n)$. A state of L_i is cataloged as the value of x_{ii} with all other x_{ij} in L_i taking on optimum values. Finding the optimum for each state becomes a subproblem which will be seen to be two classical allocation problems with bounds, one allocation problem for each arm of the L .

2.1) Mathematical Description of the Algorithm

A forward dynamic programming algorithm is defined. That is, the algorithm starts at stage 1 and works forward to stage k_1 .

From the general description of the algorithm, it can be seen that the state is defined by the index, i , of L_i . The state variable which is also the decision variable is the vector, L_i , and will be uniquely defined for each value of x_{ii} . To complete the description of the algorithm, the following must be defined:

- 1) An optimal value function which defines the maximum objective with stage i in state x with stages $i-1$ to 1 taking on values which attain a maximum objective given that stage i is in state x .
- 2) A recurrence relation which relates the optimal value function at stage i to the optimal value function at stage $i-1$.

- 3) A State equation which relates the state at stage i to the state at stage $i-1$.
- 4) A set of boundary conditions which specify the optimal value function at stage 1.

The algorithm is presented for a square matrix. It is equally applicable to any $m \times n$ matrix but the necessary notation would obscure the presentation. That is, let $m = n$.

The optimal value function is defined as

$$f_i(x) = \text{the optimal value with } L_i \text{ in state } x \\ (x_{ii} = x \text{ and the rest of } L_i \text{ is optimal}).$$

The recurrence relation is

$$f_i(x) = \max_y \{f_{i-1}(y) + g_i(x,y)\} \quad (2)$$

where $g_i(x,y) =$ the value of optimal allocation of L_i with $x_{ii} = x$, $x_{i-1,i-1} = y$ and $L_{i-1}, L_{i-2}, \dots, L_1$ taking on their respective optimal allocations. Finding $g_i(x,y)$ is the subproblem mentioned in the general description of the algorithm.

The state equations provide bounds on the value of the state vector, L_i . At stage 1, it is clear the $x_{ij} \in L_1$ must be less than or equal to row sum a_i if x_{ij} is in the column arm of L_1 or column sum b_j if x_{ij} is in the row arm of L_1 (see Figure 1 for clarification). Any $x_{ij} \in L_i$, $i > 1$ must be less than or equal to the appropriate row or column sum minus all

$x_{kl} \in \{L_{i-1}, \dots, L_1\}$ already allocated to that row or column. These bounds will be called reduced row and column sums and will be defined iteratively.

First, some additional notation is required. $\rightarrow \rightarrow 0 \ 0 \ 0 \ 0$

All quantities with a bar ($\bar{}$) on top are vectors.

Let

$\hat{y}_i(x)$ = that value of x which maximizes $f_i(x)$

$\bar{u}_i(x)$ = optimal x_{ij} , $j = 1, \dots, i-1$ with $x_{i-1,i-1} = \hat{y}_i(x)$, $x_{ii} = x$.

$\bar{v}_i(x)$ = optimal x_{ji} , $j = 1, \dots, i-1$ with $x_{i-1,i-1} = \hat{y}_i(x)$, $x_{ii} = x$.

Now, the reduced row and column sums, or state equations, may be defined

$$\bar{a}_i(x) = \{\bar{a}_{i-1} \hat{y}_i(x)\} - \bar{u}_i(x) \quad (3)$$

$$\bar{b}_i(x) = \{\bar{b}_{i-1} \hat{y}_i(x)\} - \bar{v}_i(x) \quad (4)$$

$$\bar{a}_0(0) = (a_1, a_2, \dots, a_n) \quad (5)$$

$$\bar{b}_0(0) = (b_1, b_2, \dots, b_n) \quad (6)$$

Boundary conditions, or the value of $f_1(x)$ are now easily defined.

$$f_1(x) = g_1(x, 0).$$

The subproblem of finding $g_i(x, y)$ is yet to be solved. The problem is stated mathematically as follows:

$$\begin{aligned} g_i(x, y) = & \max \sum_{j=1}^{i-1} r_{ij}(x_{ij}) \\ & + \max \sum_{j=1}^{i-1} r_{ji}(x_{ji}) \\ & + r_{ii}(x) \end{aligned} \quad (7)$$

$$\text{s.t.} \quad \sum_{j=1}^{i-1} x_{ij} = a_{i-1,i}(y) - x \quad (8)$$

$$x_{ij} \leq v_{i-1,j}(y) \quad \forall j \quad (9)$$

$$\sum_{j=1}^{i-1} x_{ji} = b_{i-1,i}(y) - x \quad (10)$$

$$x_{ji} \leq a_{i-1,j}(y) \quad \forall j \quad (11)$$

It is easily seen that $g_i(x,y)$ can be found by separating this problem into two problems of the form

$$\max \quad \sum_{i=1}^m \pi_i(z_i) \quad (12)$$

$$\text{s.t.} \quad \sum z_i = z \quad (13)$$

$$z_i \leq q_i. \quad (14)$$

By redefining $\pi_i(z_i) = -\infty$, $z_i > q_i$, and eliminating Eq. (14), the problem becomes the classical allocation problem. A backwards dynamic programming algorithm for solving the allocation problem¹ begins by letting the index i , of z_i be the stage variable and defining a state w_i to be the amount left to stages $i+1, \dots, m$. The decision variable is the amount, z_i , allocated to stage i . The optimal value function is $s_j(w_j) =$ the optimal value of an allocation for stages $j+1, \dots, m$ with w_j left before allocation of z_j . The recurrence relation is

$$s_j(w_j) = \max_{z_j} \{s_{j+1}(w_{j+1}) + \pi_j(z_j)\}. \quad (15)$$

The state equation is

$$w_{j+1} = w_j - z_j, \quad (16)$$

and the boundary conditions are

$$s_n(w_n) = \pi_n(w_n). \quad (17)$$

The solution is the policy associated with $s_j(z)$.

3) BOUNDS

It is often desirable to bound each value of x_{ij} . For example, if x_{ij} represents employees, note that in Table I missing values of employees are often listed as BB. This implies that the missing value is bounded between 250 and 499. If the lower bound is l_{ij} and the upper bound is h_{ij} , then

$$l_{ij} \leq x_{ij} \leq h_{ij}.$$

Bounds raise a problem since it is no longer necessary that a feasible solution exists. However, the problem of testing feasibility is readily solved using a max flow algorithm on a network.⁴ First, note that the added constraint can, without loss of generality, be changed to

$$0 \leq x_{ij} \leq d_{ij}$$

where $d_{ij} = h_{ij} - l_{ij}$ and a_i and b_j are reduced by l_{ij} . All of the constraints will still hold. Now define a network as shown in Figure 2 where the numbers on the arcs are upper bounds. Solve for the max flow and if the total flow = $\sum_i a_i = \sum_j b_j$ then a feasible solution exists. One feasible solution is $x_{ij} = f_{ij}$ where f_{ij} is the flow from node i to node j .

Bounds are easily incorporated into the dynamic programming algorithm. Equations (9) and (11) are changed to the following:

$$x_{ij} \leq \min \{b_{i+1,j}(y), d_{ij}\} \quad \forall j \quad (18)$$

$$x_{ji} \leq \min \{a_{i+1,j}(y), d_{ji}\} \quad \forall j. \quad (19)$$

Unfortunately, Eqs, (18) and (19) sometimes cause the algorithm to "get stuck." This occurs when, for example, for every state of stage k , too little has been allocated to the relative optimum solution for stages $\{i; i \geq k\}$ and no feasible solution exists for stage k . One can overcome this difficulty by solving the feasibility max flow problem with reduced row and column sums at each stage of the subproblem of finding $g_i(x,y)$. Little additional computation time is necessary due to the extreme efficiency of max flow algorithms.

4) AN EXAMPLE

The algorithm was used to fill in missing data in the value of shipments and value added columns of the Census of Manufactures for the state of California. For this maiden voyage, some simplifying assumptions were made. First, define the following notation:

E_C = number of employees in California in a certain industry

E_O = number of employees in the nation in that industry

S_C = number of establishments engaged in that industry

S_O = number of establishments engaged in that industry in
the nation

V_C = value added for that industry in California

V_O = value added for that industry in the nation

It is assumed that V_C is normally distributed with mean = $E_C V_O / E_O$ and variance = $(E_C V_O / E_O)^2$. When E_C is missing, it is estimated by $E_O S_C / S_O$.

Each separate maximization was disaggregated into thirty to fifty states per stage depending on the spread of maximum and minimum non-zero limits on the variables.

Table III shows a portion of the resultant estimates. The raw data is in Table I. Results seem very good. Only in a few cases were missing values replaced by zero which is a questionable value since the appearance of an industry implies that its production is non-zero.

5) CONCLUSIONS

An evaluation of the suboptimality of the algorithm would be useful. In the example provided, a quadratic programming solution could be used to find the true optimal solution and values could be compared.

One strength of the DP approach is that the reward function can be changed to any desirable shape without increasing the length of time required for solution. Experiments with more accurate and better based reward functions would be interesting.

The reduced gradient algorithm⁵ has a special appeal for this problem and deserves some investigation. As shown in Fig. 2, the problem can be constructed as a network problem. Since a basic solution forms a tree,³ each nonbasic arc is the sum of the individual gradients around the cycle. An algorithm could be derived which selectively changes the flow around these cycles and the identity of the basic variables until all cycles are either bounded above and have a negative reduced gradient, bounded below and have a positive reduced gradient or have a zero reduced gradient.

REFERENCES

1. Bellman, R.E. and Dreyfus, S.E., Applied Dynamic Programming, Princeton University Press, 1962, pp. 13-26.
2. Dantzig, G.B., Linear Programming and Extensions, Princeton University Press, 1963, pp. 35,36.
3. Ibid, pp. 356.
4. Ford, L.K., Jr. and Fulkerson, P.R., Flows in Networks, Princeton University Press, 1962, Chapter 1.
5. Leuenberger, D.G., Introduction to Linear and Non-Linear Programming, Addison Wesley, 1973, pp. 262-270.
6. U.S. Department of Commerce, Bureau of the Census, 1967 Census of Manufactures, Volume II, "Area Statistics," pp. 5-14.
7. U.S. Department of Commerce, Bureau of the Census, 1967 Commodity Transportation Survey, Part 2, "Production Areas and Selected States." Volume III.

5) CONCLUSIONS

An evaluation of the suboptimality of the algorithm would be useful. In the example provided, a quadratic programming solution could be used to find the true optimal solution and values could be compared.

One strength of the DP approach is that the reward function can be changed to any desirable shape without increasing the length of time required for solution. Experiments with more accurate and better based reward functions would be interesting.

The reduced gradient algorithm⁵ has a special appeal for this problem and deserves some investigation. As shown in Fig. 2, the problem can be constructed as a network problem. Since a basic solution forms a tree,³ each nonbasic arc is the sum of the individual gradients around the cycle. An algorithm could be derived which selectively changes the flow around these cycles and the identity of the basic variables until all cycles are either bounded above and have a negative reduced gradient, bounded below and have a positive reduced gradient or have a zero reduced gradient.

REFERENCES

1. Bellman, R.E. and Dreyfus, S.E., Applied Dynamic Programming, Princeton University Press, 1962, pp. 13-26.
2. Dantzig, G.B., Linear Programming and Extensions, Princeton University Press, 1963, pp. 35,36.
3. Ibid, pp. 356.
4. Ford, L.K., Jr. and Fulkerson, P.R., Flows in Networks, Princeton University Press, 1962, Chapter 1.
5. Leuenberger, D.G., Introduction to Linear and Non-Linear Programming, Addison Wesley, 1973, pp. 262-270.
6. U.S. Department of Commerce, Bureau of the Census, 1967 Census of Manufactures, Volume II, "Area Statistics," pp. 5-14.
7. U.S. Department of Commerce, Bureau of the Census, 1967 Commodity Transportation Survey, Part 2, "Production Areas and Selected States." Volume III.

Fig. 1. "Ls" of a 4 x 3 Matrix

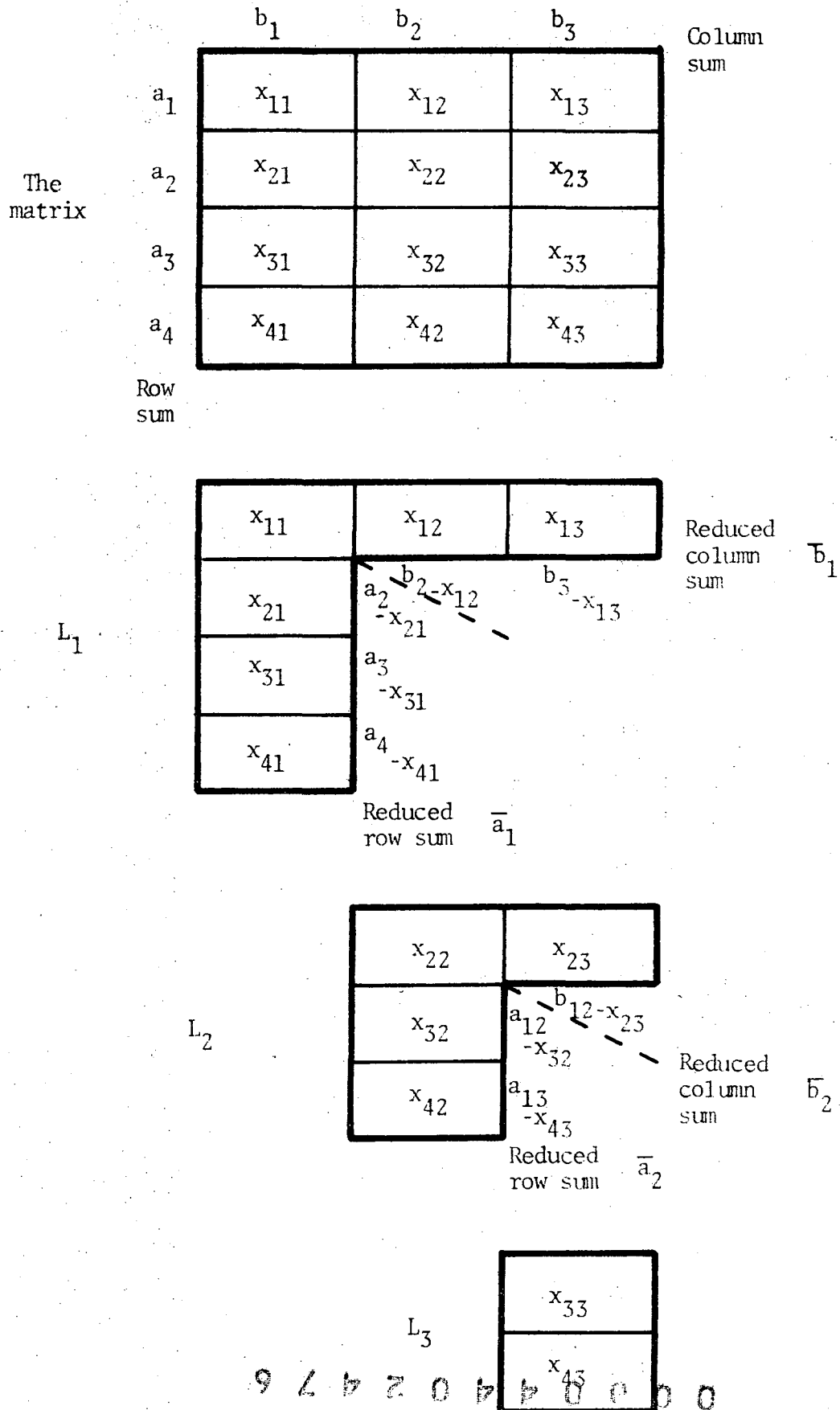
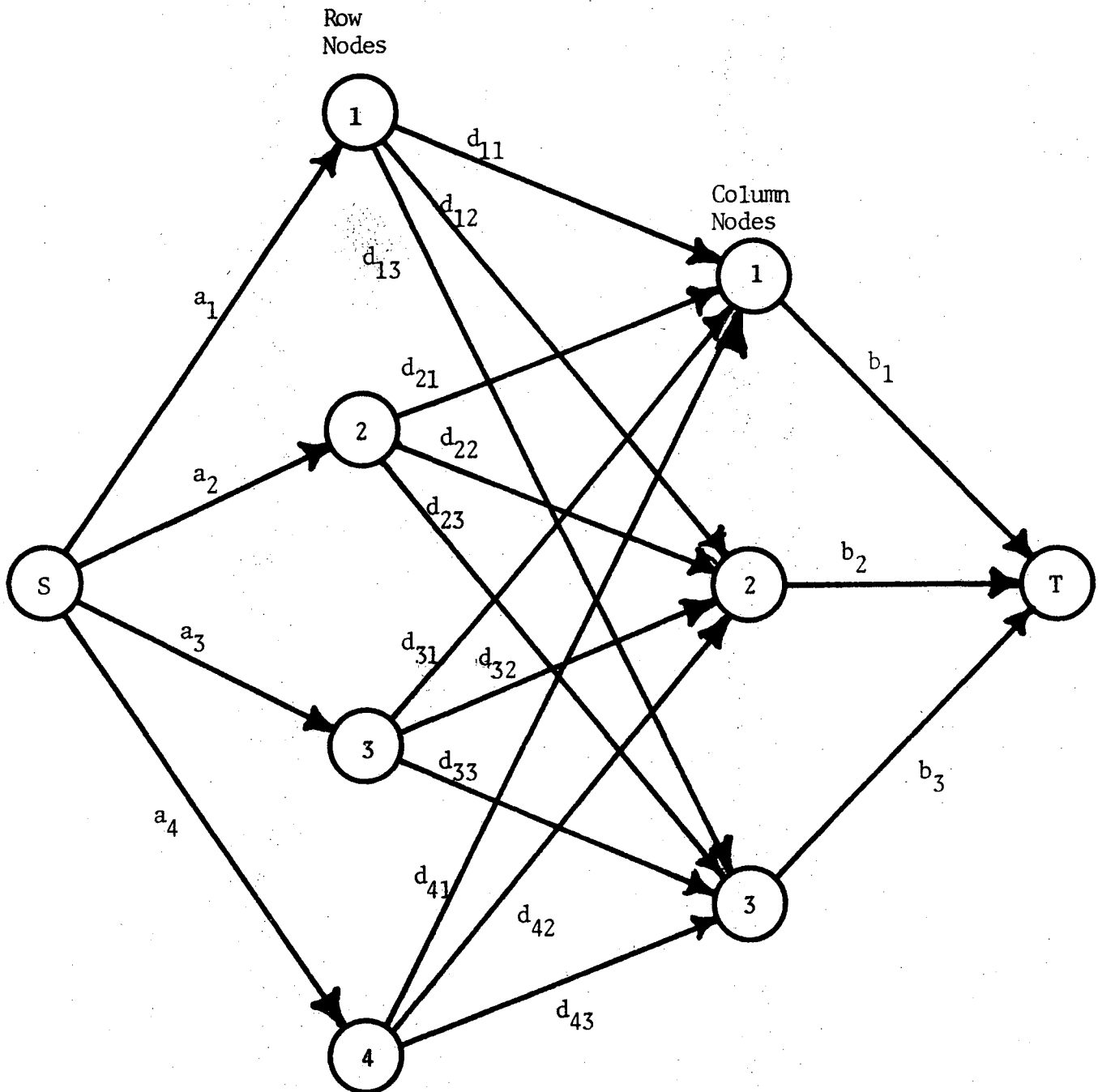


Fig. 2 Network Associated with a 4 x 3 Matrix



S = Source

T = Sink

Labels on arcs are upper bounds on flow.

TABLE I

TABLE 5. General Statistics by Industry Group and Industry: 1967 and 1963-Continued

Ind. code	Industry	1967											1963	
		Establishments		All employees		Production workers			Value added by manufacture	Cost of materials	Value of shipments	Capital expenditures, new	All employees	Value added by manufacture
		Total (number)	With 20 employees or more (number)	Number (1,000)	Payroll (million dollars)	Number (1,000)	Man-hours (millions)	Wages (million dollars)	(million dollars)	(million dollars)	(million dollars)	(million dollars)	(1,000)	(million dollars)
31	LEATHER AND LEATHER PRODUCTS . . .	197	65	FF	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	5.7	39.4
3111	LEATHER TANNING AND FINISHING	15	8	7	4.7	6	1.2	3.7	8.4	11.5	19.5	(0)	6	4.4
3121	INDUSTRIAL LEATHER BELTING	6	2	BB	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3131	FOOTWEAR CUT STOCK	7	3	AA	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	1	9
314	FOOTWEAR, EXCEPT RUBBER	44	19	2.4	11.2	2.1	3.9	8.1	18.1	15.0	32.9	2	2.5	15.8
3141	SHOES, EXCEPT RUBBER	42	19	EE	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3151	LEATHER GLOVES AND MITTENS	9	4	2	6	1	2	5	1.0	9	1.9	(2)	2	9
3161	LUGGAGE	28	9	CC	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
317	HANDBAGS, PERSONAL LEATHER GOODS . .	46	14	1.2	6.0	1.0	1.9	4.3	9.8	6.4	16.1	(0)	1.0	6.1
3171	WOMEN'S HANDBAGS AND PURSES	22	11	CC	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3172	PERSONAL LEATHER GOODS	24	3	BB	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3199	LEATHER GOODS, NEC	42	6	4	1.9	8	7	1.3	3.2	3.4	6.6	(0)	1.3	2.0
32	STONE, CLAY, AND GLASS PRODUCTS . .	1 435	409	43.4	306.9	33.3	66.1	216.1	662.6	519.8	1 190.9	63.0	45.6	654.0
3211	FLAT GLASS	10	5	EE	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
322	GLASS, GLASSWARE, PRESSED OR BLOWN	26	19	6.2	57.3	7.2	14.3	48.7	110.2	62.6	178.5	(0)	(0)	(0)
3221	GLASS CONTAINERS	18	16	FF	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3229	PRESSED AND BLOWN GLASS, NEC	10	3	BB	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	4	4.8
3231	PRODUCTS OF PURCHASED GLASS*	119	27	1.6	9.7	1.3	2.5	6.5	19.3	16.4	35.7	(0)	1.6	20.8
3241	CEMENT, HYDRAULIC	14	13	3.6	30.1	2.8	5.7	21.4	95.4	46.1	141.9	7.9	4.3	93.1
325	STRUCTURAL CLAY PRODUCTS	70	38	4.1	26.2	3.2	6.5	18.5	46.9	22.5	70.3	2.8	5.2	59.7
3251	BRICK AND STRUCTURAL CLAY TILE	27	13	7	4.8	6	1.2	3.6	6.1	3.0	9.8	5	1.0	8.4
3253	CERAMIC WALL AND FLOOR TILE	19	12	2.0	13.5	1.5	3.2	8.5	20.6	9.2	29.6	(0)	2.4	23.8
3255	CLAY REFRACTORIES	10	5	BB	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3259	STRUCTURAL CLAY PRODUCTS, NEC	14	8	EE	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	1.7	23.9
326	POTTERY AND RELATED PRODUCTS	158	42	FF	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	4.0	32.1
3261	VITREOUS PLUMBING FIXTURES	10	7	1.1	7.9	9	1.7	6.2	16.6	9.0	25.6	(0)	1.1	14.8
3262	VITREOUS CHINA FOOD UTENSILS	8	6	3	1.3	3	6	1.2	1.9	4	2.2	(0)	(0)	(0)
3263	FINE EARTHENWARE FOOD UTENSILS	7	5	CC	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3264	PORCELAIN ELECTRICAL SUPPLIES	11	2	AA	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3269	POTTERY PRODUCTS, NEC	122	22	1.8	7.5	1.8	3.0	6.1	13.3	4.3	17.5	8	1.9	11.2
327	CONCRETE, GYPSUM, PLASTER PRODUCTS	787	198	13.2	101.7	8.7	16.9	62.0	218.6	245.7	464.6	18.7	15.5	232.8
3271	CONCRETE BLOCK AND BRICK	81	19	1.1	7.7	7	1.4	4.5	14.6	12.1	26.8	1.2	1.2	16.6
3272	CONCRETE PRODUCTS, NEC	237	50	4.9	37.0	3.7	7.3	25.6	74.0	48.9	125.5	5.8	4.9	57.6
3273	READY-MIXED CONCRETE	442	108	6.2	49.1	3.4	8.4	26.0	106.7	165.3	271.7	9.9	7.8	124.3
3274	LIME	6	8	BB	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3275	GYPSUM PRODUCTS	21	9	CC	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3281	CUT STONE AND STONE PRODUCTS*	51	7	5	3.3	4	8	2.4	5.3	3.1	8.5	2	6	4.5
329	MISC. NONMETALLIC MINERAL PRODUCTS	200	60	6.9	49.3	5.3	10.7	33.9	110.5	94.4	206.1	13.3	6.9	119.0
3291	ABRASIVE PRODUCTS	26	6	5	3.5	3	7	2.0	8.7	5.4	13.9	3	(0)	(0)
3292	ASBESTOS PRODUCTS	18	16	EE	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	1.9	81.3
3293	GASKETS AND INSULATIONS	30	13	1.5	10.5	1.1	2.3	6.2	16.2	10.6	26.7	(0)	1.1	11.3
3295	MINERALS, GROUND OR TREATED	53	11	9	6.2	7	1.3	4.4	15.2	11.9	27.0	5.8	9	14.9
3296	MINERAL WOOL	12	6	EE	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3297	NONCLAY REFRACTORIES	6	2	BB	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3299	NONMETALLIC MINERAL PRODUCTS, NEC . .	55	6	4	2.4	3	6	1.5	3.7	4.7	8.4	(0)	4	5.3
33	PRIMARY METAL INDUSTRIES	742	377	58.8	456.2	47.3	96.1	334.5	933.3	1 201.7	2 103.0	92.4	48.5	603.9
331	BLAST FURNACE, BASIC STEEL PROD. . . .	84	47	22.4	191.3	17	35.0	140.7	378.8	562.0	907.6	32.5	19.5	257.4
3312	BLAST FURNACES AND STEEL MILLS	30	18	FF	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	16.2	231.7
3315	STEEL WIRE AND RELATED PRODUCTS	23	12	1.4	11.0	1.1	2.2	7.7	23.2	29.3	51.9	(0)	(0)	(0)
3316	COLD FINISHING OF STEEL SHAPES	7	3	AA	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3317	STEEL PIPE AND TUBES	22	13	FF	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	1.8	19.3
332	IRON AND STEEL FOUNDRIES	101	76	8.0	53.3	7.0	13.5	41.8	94.3	55.2	150.2	6.7	7.4	73.9
3321	GRAY IRON FOUNDRIES	63	48	4.2	27.0	3.6	7.0	21.4	44.0	32.8	78.5	2.4	4.3	46.4
3323	STEEL FOUNDRIES	35	29	FF	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
333	PRIMARY NONFERROUS METALS	7	3	BB	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3332	PRIMARY LEAD	3	3	BB	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3341	SECONDARY NONFERROUS METALS	45	21	1.6	11.5	1.1	2.2	6.9	27.8	82.8	112.3	2.7	1.5	20.0
335	NONFERROUS ROLLING AND DRAWING	83	54	FF	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
3351	COPPER ROLLING AND DRAWING	6	5	EE	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	1.3	18.8
3352	ALUMINUM ROLLING AND DRAWING	29	17	7.6	61.2	6.0	13.6	46.0	132.2	161.9	296.6	17.0	5.2	70.9
3356	NONFERROUS ROLLING AND DRAWING, NEC*	13	4	4	3.1	4	7	2.3	9.3	15.5	23.8	9	(0)	(0)
3357	NONFERROUS WIREDRAWING, INSULATING . .	35	28	3.3	24.0	2.5	5.3	15.9	69.0	102.7	166.7	(0)	(0)	(0)
336	NONFERROUS FOUNDRIES	243	100	7.6	50.2	6.5	13.0	38.3	89.3	65.2	153.8	5.3	6.7	67.3
3361	ALUMINUM CASTINGS	139	60	4.2	27.9	3.5	7.2	21.3	49.2	30.6	79.2	(0)	3.5	32.9
3362	BRASS, BRONZE, AND COPPER CASTINGS . .	57	19	1.1	7.4	1.0	1.9	5.5	14.0	29.7	44.8	(0)	1.0	9.7
3369	NONFERROUS CASTINGS, NEC	47	21	2.4	14.8	2.1	4.0	11.5	26.1	18.7	44.8	(0)	2.2	24.6
339	MISC. PRIMARY METAL PRODUCTS	179	76	6.3	49.7	4.9	10.4	34.4	106.2	73.2	181.7	15.1	3.6	48.0
3391	IRON AND STEEL FORGINGS	26	12	1.4	12.4	1.1	2.2	8.6	26.0	31.2	57.4	(0)	1.1	17.0
3392	NONFERROUS FORGINGS	11	10	1.5	12.7	1.1	2.4	8.7	25.5	26.0	51.9	6.9	9	9.7
3399	PRIMARY METAL PRODUCTS, NEC	142	54	3.4	24.6	2.7	5.7	17.0	54.7	18.1	72.3	(0)	1.7	21.3

See footnotes at end of table.

8 7 6 2 0 4 0 0 0 0

TABLE II

36

GENERAL SUMMARY

TABLE 3. General Statistics for Establishments by Industry Groups and Industries:
1967 and 1963—Continued

1967											
Code	Industry group and industry	Companies (number)	Establishments		All employees		Production workers			Value added by manufacture (million dollars)	Value of shipments (million dollars)
			Total (number)	With 20 employees or more (number)	Number (1,000)	Payroll (million dollars)	Number (1,000)	Man-hours (millions)	Wages (million dollars)		
32	STONE, CLAY, AND GLASS PRODUCTS--CON.										
326	POTTERY AND RELATED PRODUCTS	574	619	231	42.7	246.9	36.0	70.1	190.0	418.7	*605.4
3261	VITREOUS PLUMBING FIXTURES	34	31	40	8.4	56.5	7.0	13.8	45.7	111.4	170.0
3262	VITREOUS CHINA FOOD UTENSILS	33	39	28	7.0	36.1	6.0	11.2	27.3	55.8	67.2
3263	FINE EARTHENWARE FOOD UTENSILS	20	21	17	5.9	27.5	5.3	9.8	22.6	34.6	47.1
3264	PORCELAIN ELECTRICAL SUPPLIES	63	74	50	13.4	89.9	11.0	22.2	67.5	153.5	225.0
3269	POTTERY PRODUCTS, NEC.	424	434	96	7.9	37.0	6.7	13.1	27.0	63.4	96.1
327	CONCRETE, GYPSUM, PLASTER PRODUCTS	8 638	10 004	2 450	169.1	1 103.3	122.7	255.0	719.0	2 478.0	*4 999.6
3271	CONCRETE BLOCK AND BRICK	1 534	1 599	349	21.1	130.1	14.6	30.6	77.0	274.2	550.1
3272	CONCRETE PRODUCTS, NEC.	3 055	3 412	802	56.8	351.1	45.4	94.2	247.9	722.6	1 201.1
3273	READY-MIXED CONCRETE	4 032	4 760	1 150	74.5	508.8	49.2	100.7	308.6	1 155.5	2 684.2
3274	LIME	78	116	71	6.8	42.6	5.7	12.0	33.4	100.1	176.5
3275	GYPSUM PRODUCTS	55	117	78	10.0	70.7	7.8	17.5	52.2	225.6	387.7
3281	CUT STONE AND STONE PRODUCTS	776	811	196	16.6	90.8	14.1	28.3	70.9	148.2	*236.4
329	MISC. NONMETALLIC MINERAL PRODUCTS	1 485	1 842	708	106.1	725.8	80.7	162.1	499.7	1 546.5	2 795.3
3291	ABRASIVE PRODUCTS	320	361	165	27.2	197.5	18.9	36.3	119.2	420.7	725.3
3292	ASBESTOS PRODUCTS	81	138	99	21.3	144.0	16.8	35.0	107.5	308.1	575.0
3293	GASKETS AND INSULATIONS	288	301	123	18.5	115.3	14.1	27.8	76.8	205.5	352.8
3295	MINERALS, GROUND OR TREATED	382	504	130	8.9	59.3	7.0	14.4	41.7	149.4	280.9
3296	MINERAL WOOL	77	116	74	15.1	106.9	12.2	25.5	83.9	247.3	454.4
3297	NONCLAY REFRACTORIES	59	92	64	9.9	72.1	7.4	14.8	49.5	158.7	307.1
3299	NONMETALLIC MINERAL PRODUCTS, NEC.	325	330	53	5.3	30.6	4.3	8.3	21.1	56.8	99.9
33	PRIMARY METAL INDUSTRIES	5 538	6 837	4 082	1 281.0	9 850.9	1 041.5	2 088.7	7 457.3	19 978.2	*46 730.9
331	BLAST FURNACE, BASIC STEEL PROD.	563	861	673	617.3	5 021.9	500.8	981.6	3 829.2	10 170.1	23 121.0
3312	BLAST FURNACES AND STEEL MILLS	200	329	259	533.1	4 385.3	434.0	845.4	3 368.4	8 910.1	19 620.6
3313	ELECTROMETALLURGICAL PRODUCTS	21	34	31	10.4	79.0	8.4	17.2	58.6	193.2	467.9
3315	STEEL WIRE AND RELATED PRODUCTS	200	240	171	19.5	194.8	21.7	44.1	158.0	266.1	845.2
3316	COLD FINISHING OF STEEL SHAPES	77	107	86	19.5	154.9	14.9	29.8	108.8	269.1	1 038.7
3317	STEEL PIPE AND TUBES	123	151	126	27.0	207.9	21.8	45.1	155.4	412.1	1 148.6
332	IRON AND STEEL FOUNDRIES	1 265	1 438	1 097	232.6	1 640.5	198.3	402.1	1 306.7	2 631.0	*4 289.3
3321	GRAY IRON FOUNDRIES	969	1 061	774	138.0	965.3	119.3	242.3	780.2	1 543.1	2 637.8
3322	MALLEABLE IRON FOUNDRIES	72	81	75	25.4	187.0	21.4	43.0	147.3	256.5	438.3
3323	STEEL FOUNDRIES	256	296	248	69.2	488.2	57.6	116.8	379.2	791.4	1 213.2
333	PRIMARY NONFERROUS METALS	77	144	121	53.4	405.1	42.2	85.4	300.5	1 381.5	*3 714.0
3331	PRIMARY COPPER	11	32	32	11.6	80.6	9.2	18.1	61.0	262.6	1 184.1
3332	PRIMARY LEAD	10	19	18	2.7	18.9	2.1	4.3	13.3	48.3	304.0
3333	PRIMARY ZINC	11	18	18	8.1	57.8	6.4	13.5	42.9	119.5	332.8
3334	PRIMARY ALUMINUM	8	25	24	23.8	190.9	19.2	38.4	145.2	811.8	1 608.7
3339	PRIMARY NONFERROUS METALS, NEC.	41	50	29	7.2	56.9	5.2	11.2	38.1	139.5	284.4
3341	SECONDARY NONFERROUS METALS	351	403	182	17.2	122.8	12.8	26.6	78.4	271.2	1 586.0
335	NONFERROUS ROLLING AND DRAWING	527	833	620	194.9	1 438.4	151.1	313.4	1 027.0	3 324.9	*9 925.4
3351	COPPER ROLLING AND DRAWING	86	125	108	40.3	293.9	32.2	64.5	219.0	704.4	2 391.1
3352	ALUMINUM ROLLING AND DRAWING	155	218	163	61.5	479.4	48.7	101.0	354.9	938.7	2 958.9
3356	NONFERROUS ROLLING AND DRAWING, NEC.	115	142	74	21.3	163.5	14.8	30.9	100.4	351.7	984.0
3357	NONFERROUS WIREDRAWING, INSULATING	206	348	275	71.7	501.6	55.3	117.0	352.7	1 330.1	3 591.4
336	NONFERROUS FOUNDRIES	1 827	1 886	798	89.6	609.3	76.0	154.5	464.0	1 067.5	*1 922.0
3361	ALUMINUM CASTINGS	970	992	411	45.2	312.9	38.8	78.8	243.6	516.6	920.5
3362	BRASS, BRONZE, AND COPPER CASTINGS	516	534	207	17.9	119.8	15.1	30.6	89.2	226.8	445.2
3369	NONFERROUS CASTINGS, NEC.	352	360	180	26.6	176.7	22.1	45.0	131.2	324.1	556.3
339	MISC. PRIMARY METAL PRODUCTS	1 140	1 272	591	76.0	612.9	60.5	125.0	451.5	1 131.9	*2 173.2
3391	IRON AND STEEL FORGINGS	248	272	189	41.2	345.9	33.4	67.2	265.6	607.4	1 261.6
3392	NONFERROUS FORGINGS	37	41	31	10.1	87.5	7.4	16.5	61.4	154.4	333.1
3399	PRIMARY METAL PRODUCTS, NEC.	868	959	371	24.7	179.4	19.7	41.4	124.5	370.2	578.5
34	FABRICATED METAL PRODUCTS	25 262	27 418	10 742	1 341.8	9 319.5	1 056.9	2 160.9	6 541.6	18 042.6	*34 577.8
3411	METAL CANS	96	299	241	60.3	474.4	52.3	112.7	393.3	1 141.5	2 890.6
342	CUTLERY, HANDTOOLS, AND HARDWARE	1 810	1 939	654	155.9	1 043.1	123.9	259.9	742.2	2 318.3	*3 714.6
3421	CUTLERY	138	146	69	13.0	80.2	10.4	20.5	54.5	276.3	378.5
3423	HAND AND EDGE TOOLS, NEC.	630	667	252	35.9	232.9	28.7	57.9	165.8	509.4	813.5
3425	HANDSAWS AND SAW BLADES	74	85	49	6.3	43.4	4.8	10.1	29.5	97.1	154.3
3429	HARDWARE, NEC.	979	1 041	484	100.8	686.7	80.0	171.4	492.4	1 435.5	2 368.3
343	PLUMBING AND HEATING EXC. ELECTRIC	781	838	389	65.2	427.9	48.5	96.8	278.4	871.9	*1 724.9
3431	METAL SANITARY WARE	83	91	51	10.1	71.3	8.1	16.4	53.5	140.6	251.2
3432	PLUMBING FITTINGS AND BRASS GOODS	210	214	90	14.5	92.6	11.4	22.8	64.2	183.5	404.7
3433	HEATING EQUIPMENT, EXCEPT ELECTRIC	496	533	248	40.6	264.0	29.0	57.5	160.7	547.9	*1 069.0
344	FABRICATED STRUCTURAL METAL PROD.	9 558	10 151	3 705	387.9	2 685.8	292.5	591.5	1 777.9	4 934.3	*10 440.2
3441	FABRICATED STRUCTURAL STEEL	1 865	1 960	1 035	107.5	773.6	80.1	163.6	509.9	1 519.3	2 967.8
3442	METAL DOORS, SASH, AND TRIM	1 540	1 646	597	63.9	378.5	47.9	94.8	235.7	651.1	1 413.6
3443	FABRICATED PLATEWORK (BOILER SHOPS)	1 346	1 465	784	96.4	700.4	73.1	150.4	477.5	1 393.4	2 688.5
3444	SHEET METALWORK	2 942	3 054	811	68.1	490.7	52.0	105.8	336.0	889.3	1 812.4
3446	ARCHITECTURAL METALWORK	1 360	1 377	198	19.3	125.0	15.4	29.5	81.5	232.6	404.9
3449	MISCELLANEOUS METALWORK	594	649	280	32.7	220.7	23.2	47.4	137.4	448.6	1 082.0
345	SCREW MACHINE PRODUCTS, BOLTS, ETC.	2 415	2 536	1 073	116.1	847.1	93.7	196.7	605.7	1 620.9	*2 665.4
3451	SCREW MACHINE PRODUCTS	1 847	1 874	681	49.0	349.1	41.5	88.2	258.5	651.5	1 021.5
3452	BOLTS, NUTS, RIVETS, AND WASHERS	577	662	292	67.2	498.1	52.2	108.5	347.2	979.4	1 643.9
3461	METAL STAMPINGS	2 564	2 710	1 351	225.9	1 685.5	187.0	379.8	1 272.4	3 030.8	5 893.7
347	MISC. FABRICATED METAL PRODUCTS	1 101	1 101	1 101	41.3	441.4	69.0	137.4	356.1	864.3	*1 262.4

TABLE III

CALIFORNIA DATA - CENSUS OF MANUFACTURERS. DISCLOSURES ESTIMATED USING DP ALGORITHM				
SIC CODE	NUMBER OF ESTABS	NUMBER OF EMPLOYS	VALUE ADDED	VALUE OF SHIPMENTS
2844	89	39.0	1271.0	1837.0
2850	235	72.0	1499.0	3498.0
2851	235	72.0	1499.0	3498.0
2870	92	28.0	499.0	1590.0
2871	5	5.0	127.0	287.0
2872	40	7.4	127.0	375.0
2879	47	16.0	244.0	927.0
2890	243	44.0	856.0	1670.0
2891	33	7.0	143.0	308.0
2892	10	5.0	55.0	114.0
2893	46	12.0	145.0	426.0
2899	151	21.0	512.0	822.0
2900	201	164.0	5542.0	23546.0
2910	50	139.0	4966.0	22127.0
2911	50	139.0	4966.0	22127.0
2950	111	18.0	368.0	969.0
2951	90	8.0	221.0	532.0
2952	21	10.0	147.0	437.0
2990	40	6.0	208.0	450.0
2992	37	9.1	207.0	450.0
3000	870	372.0	5230.0	9724.0
3010	28	142.6	1404.0	2981.0
3011	28	142.6	1404.0	2981.0
3020	2	9.0	48.0	0.0
3021	2	9.0	48.0	0.0
3060	146	75.0	993.0	1744.0
3069	146	75.0	993.0	1744.0
3070	691	218.0	2784.0	4998.0
3079	691	218.0	2784.0	4998.0
3100	197	175.7	1226.0	3272.0
3110	15	7.0	84.0	195.0
3111	15	7.0	84.0	195.0
3120	6	2.5	109.0	250.0
3121	6	2.5	109.0	250.0
3130	7	2.5	27.0	166.0
3131	7	2.5	27.0	166.0
3140	44	24.0	181.0	329.0
3141	42	25.0	181.0	329.0
3150	9	2.0	10.0	19.0
3151	9	2.0	10.0	19.0
3160	28	10.0	683.0	2084.0
3161	28	10.0	683.0	2084.0
3170	46	12.0	98.0	161.0
3171	22	10.0	62.0	101.0
3172	24	5.0	35.0	59.0
3190	42	4.0	32.0	66.0
3199	42	4.0	32.0	66.0
3200	1435	434.0	6626.0	11909.0
3210	10	25.0	244.0	399.0
3211	10	25.0	244.0	399.0
3220	26	82.0	1102.0	1785.0
3221	16	88.9	1064.0	1725.0
3229	10	5.0	36.0	59.0

6 1 2 0 7 0 0 0 0

TABLE III (Continued)

CALIFORNIA DATA - CENSUS OF MANUFACTURERS. DISCLOSURES ESTIMATED USING DP ALGORITHM				
SIC CODE	NUMBER OF ESTABS	NUMBER OF EMPLOYS	VALUE ADDED	VALUE OF SHIPMENTS
3230	119	16.0	193.0	357.0
3231	119	16.0	193.0	357.0
3240	14	36.0	954.0	1419.0
3241	14	36.0	954.0	1419.0
3250	70	41.0	469.0	703.0
3251	27	7.0	61.0	98.0
3253	19	20.0	206.0	296.0
3255	10	5.0	67.0	113.0
3259	14	10.0	134.0	195.0
3260	158	109.0	318.0	453.0
3261	10	11.0	166.0	256.0
3262	8	3.0	19.0	22.0
3263	7	10.0	0.0	0.0
3264	11	2.5	0.0	0.0
3269	122	18.0	133.0	175.0
3270	787	132.0	2186.0	4646.0
3271	81	11.0	146.0	268.0
3272	237	49.0	760.0	1255.0
3273	442	62.0	1067.0	2717.0
3274	6	3.5	49.0	94.0
3275	21	10.0	163.0	311.0
3280	51	5.0	53.0	85.0
3281	51	5.0	53.0	85.0
3290	200	69.0	1105.0	2061.0
3291	26	5.0	87.0	139.0
3292	18	25.0	355.0	737.0
3293	30	15.0	162.0	267.0
3295	53	9.0	152.0	270.0
3296	12	15.6	244.0	433.0
3297	6	5.0	66.0	130.0
3299	55	4.0	37.0	84.0
3300	742	588.0	9333.0	21030.0
3310	84	224.0	3788.0	9076.0
3312	30	486.1	3081.0	7129.0
3315	23	14.0	232.0	519.0
3316	7	2.5	0.0	0.0
3317	22	39.3	474.0	1425.0
3320	101	80.0	943.0	1502.0
3321	63	42.0	440.0	785.0
3323	35	81.8	503.0	716.0
3330	7	5.0	131.0	324.0
3332	3	4.3	131.0	324.0
3340	45	16.0	278.0	1123.0
3341	45	16.0	278.0	1123.0
3350	83	194.2	2236.0	5649.0
3351	6	19.3	130.0	757.0
3352	29	76.0	1322.0	2966.0
3356	13	4.0	93.0	238.0
3357	35	33.0	690.0	1687.0
3360	243	76.0	893.0	1538.0
3361	139	42.0	492.0	792.0
3362	57	11.0	140.0	297.0
3369	47	24.0	261.0	448.0

LEGAL NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Energy Research and Development Administration, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

TECHNICAL INFORMATION DIVISION
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720