Economic Classification Policy Committee

Report No. 2

The Heterogeneity Index: A Quantitative Tool to Support Industrial Classification

August 1994

This report was prepared by Frank M. Gollop, Department of Economics, Boston College, and is based on theoretical work by Gollop and empirical computations by Ron Jarmin, of the Center for Economic Studies, Bureau of the Census, U.S. Department of Commerce. Data for the report are from the Census Bureau's Longitudinal Research Database.

Copies of this and other Economic Classification Policy Committee reports can be obtained from the Economic Classification Policy Committee, Bureau of Economic Analysis (BE-42), U.S. Department of Commerce, Washington, D.C. 20230, or by telephone at (202) 606-9615, FAX (202) 606-5311.

Economic Classification Policy Committee

Report No. 2

The Heterogeneity Index: A Quantitative Tool to Support Industrial Classification

The three North American nations have been working jointly to establish a common North American system of industrial classifications. After evaluating alternative conceptual bases, the Economic Classification Policy Committee (ECPC) in the United States, the Mexican Instituto Nacional de Estadística, Geografía e Informática, and Statistics Canada have adopted the position that industrial classifications in the North American Industry Classification System (NAICS) should conform to the "production-oriented," or "supply-based," concept.[1] Establishments should be grouped into industries based on similar production processes, or in the language of economics, similar production functions.

Separate analyses of the current industrial classifications in the United States and Canada reveal that neither country's system conforms to a single conceptual basis but instead represents a mix of production and demand-based concepts.[2] One objective of the multicountry effort is to move each country's industrial classification to a consistent production-based system.

There is little doubt that informed judgment based on, among other things, engineering evidence and institutional knowledge will be the ultimate arbiter in identifying proper classes of economic activity and in assigning establishments to those industrial classes. Much of this process will, by necessity, be qualitative and judgmental in nature. However, just as a medical diagnosis is aided significantly by quantitative tools like the simple thermometer, the process of industrial classification could be greatly enhanced by the availability of measurements capable of quantifying the homogeneity in each industry grouping.

---

[1] See the joint statement on the concept for NAICS in *Federal Register*, July 26, 1994, pp. 30892-30896 (Part II). ECPC Issues Paper No. 1, "Conceptual Issues," discusses alternative classification concepts, including the market-oriented, or demand-based concept.

[2] See ECPC Report No. 1, "Economic Concepts Incorporated in the Standard Industrial Classification Industries of the United States," July 1994, and "The Conceptual Basis of the Standard Industrial Classification," by Kenneth Young, Statistics Canada, February 1994.

This paper presents and discusses an analytic measurement--the heterogeneity index--that can serve as a quantitative complement to the tools already available for designing and maintaining an industrial classification system that is based on the production-oriented concept.

Section 1 derives and discusses the new measure, a variant of the heterogeneity component of the diversification index introduced in Gollop and Monahan (1991).[3] In brief, the new heterogeneity index quantifies the extent of similarity among the production functions represented by the establishments assigned to an industry category. Relying on U.S. data, section 2 offers evidence supporting the index's application to the process of industrial classification. Section 3 suggests a variety of specific practical uses for the index in developing and maintaining an industrial classification system. Section 4 discusses possibilities for the index's enhancement, and section 5 concludes.

1. The Heterogeneity Index

A "production-oriented" concept for industrial classification establishes the criterion that those establishments having similar production processes should be grouped together in a common industrial category while those exhibiting dissimilar production processes should be assigned to different industries. In the economic theory of production, an establishment's entire set of production relationships is summarized in its production function, which relates inputs to each other and to output. A statistical measure suitable to the task of testing or defining appropriate boundaries for an industry must discern the extent of heterogeneity among the production functions belonging to the incumbent or candidate establishments in a particular industry.

The properties of a production function are captured in parameters defining the relationships among inputs and outputs. Identical production function parameters across establishments suggest homogeneous technologies while different parameters specify heterogeneous technologies. Identifying these parameters is the key to designing a statistical measure that can assist industrial classification.

It turns out that, under reasonable assumptions, the information required for identifying these production parameters can be extracted from data commonly available in industrial accounts. To demonstrate this, consider one of the simplest of

[3] Frank M. Gollop and James L. Monahan, "A Generalized Index of Diversification: Trends in U.S. Manufacturing," *The Review of Economics and Statistics*, LXIII (May), pp. 318-30.

economic production functions, the Cobb-Douglas production function:[4]

$$(1)\quad y_i = \Pi_j X_{ij}^{\beta_{ij}},$$

where y_i is a vector of outputs produced by the ith establishment using a set of inputs, X_j, and a Cobb-Douglas technology described by the parameters β_{ij} (and where Π_j indicates the product of the input terms). Assuming competitive input markets, the Cobb-Douglas parameters β_{ij} associated with the inputs are equal to the corresponding input cost shares

$$(2)\quad \beta_{ij} = \frac{\partial lny_i}{\partial lnX_{ij}} = \frac{p_{ij}X_{ij}}{\Sigma_j p_{ij}X_{ij}} = w_{ij},$$

where p_{ij} is the price of the jth input, X_{ij} is the quantity of it used in production, and w_{ij} is the cost share of the jth input in the total input costs of the ith establishment.

If we consider another establishment, k, which also uses a Cobb-Douglas technology, this technology will correspond to parameters B_{kj}, and accordingly to input cost shares w_{hj}. If the two establishments have the same technology, then $B_{kj} = B_{ij}$. But if this is so, then we know from equation (2) that the input cost shares of the two establishments will also be the same, that is: $w_{kj} = w_{ij}$. If the production parameters in the two establishments are <u>not</u> the same--that is, if $B_{kj} \neq B_{ijm}$--then it will also be true that the input cost shares will differ, that is, $w_{kj} \neq w_{ij}$. Differences in input cost shares among establishments therefore quantify differences among production parameters.

Differences among production parameters across establishments, in turn, can be used to calibrate the extent of heterogeneity among the establishments' production functions. A production-based index of heterogeneity, H, for establishments within an industry follows directly:

$$(3)\quad H = \Sigma_i \Sigma_k\, s_i s_k \,\frac{(\Sigma_j |w_{ij} - w_{kj}|)}{2},$$

where s_i and s_k are the respective shares of the ith and kth establishments in industry sales, and w_{ij} and w_{kj} are the input

---

[4] For an introduction to production functions, see Walter Nicholson, *Microeconomic Theory: Basic Principles and Extensions*. Hilsdale, Illinois: Dryden Press (1972), chapter 11, or Hal R. Varian, *Microeconomic Analysis* (2nd edition), New York and London: W. W. Norton and Company (1984), chapter 4.

cost shares of the jth input in the ith and kth establishments, respectively. Division by 2 prevents double counting and ensures that the index H is bounded in the zero-one interval, $0 \leq H < 1$.

The heterogeneity index H defined in (3) is simply a weighted average over differences in production parameters describing the technologies employed in establishments within an industry. As differences among those parameters increase, H increases; as the differences decrease, the index H approaches zero.

Note that the establishment shares s_i and s_k in (3) play an important role in the definition of H. For any given difference in the input shares of the ith and kth establishments, the overall effect on industry H is determined by the relative importance of the ith and kth establishments. Input differences between large establishments have more impact on H than do input differences between small establishments. The share variables s_i and s_k ensure this result.

It is instructive to rewrite (3) in its equivalent form

$$(4) \quad H = \Sigma_i \, s_i H_i \,,$$

$$\text{where } H_i = \Sigma_k \, s_k \, \frac{(\Sigma_j |w_{ij} - w_{kj}|)}{2} \,.$$

The variable H_i quantifies the difference between the production function of the ith establishment and the production functions of all other establishments in the industry. The product s_iH_i identifies the contribution of the ith establishment to industry-wide heterogeneity H. The contribution of each establishment to industry-wide H depends on both the extent of the establishment's heterogeneity with respect to other establishments in the industry and the establishment's share in industry sales--differences in production parameters among large establishments make a greater contribution to the industry heterogeneity index than do similar differences in production parameters among small establishments.

Applications of the heterogeneity index are discussed in full in section 3 below. One application, however, follows directly from equation (4) and merits mention here. In those industries where H is found to be large, overall H can be decomposed using equation (4) into establishment-specific heterogeneity indexes H_i. The "offending" establishments can be identified and the effect of their heterogeneity s_iH_i can be quantified.

2. Evidence from U.S. Data

The heterogeneity index defined in equation (3) was constructed for 175 4-digit manufacturing industries as defined by the 1987 U.S. Standard Industrial Classification. The industries were those chosen, independently, for review by a team assembled by the ECPC which produced the results reported in ECPC Report No. 1.

Establishment-specific data were drawn from the 1987 Census of Manufactures. The index for each industry is based on vectors of input shares constructed for each establishment in that industry for the following inputs: production workers, other labor, fuel, electricity, purchased services, agricultural materials, mineral inputs, nondurable materials, durable materials, and capital. It is important to note that the indexes are calculated using the full population of establishments in each industry.[5]

The indexes for the 175 industries are transformed into percentile form. The lowest value of the heterogeneity index among the 175 industries takes a 0 percentile ranking. The highest value takes a value equal to 100. Intermediate index values are then scaled between 0 and 100.

The percentiles are then combined with the supply-based analysis found in the industry-classification matrix prepared by a research team under the direction of the ECPC. That matrix is presented in full as an Appendix to ECPC Report No. 1.

It is not necessary for the immediate purposes of this paper to describe the detailed steps in the "supply-based" analysis underlying the original ECPC matrix. That is done thoroughly in ECPC Report No. 1.[6] It is sufficient to state that a supply-based, or production-oriented, industry is one which the ECPC team judged to be uniquely defined in terms of the production process itself, the materials used in the production process, and/or the type of labor employed in the industry. Column entries identify which one or more (if any) supply-based criteria define a particular industry. Blanks in all columns for an industry indicate that the ECPC team concluded that the industry's current configuration of establishments is not consistent with any supply-based criteria.

Before evaluating the extent of any correspondence between the calculated heterogeneity index and the ECPC's supply-based analysis, it is important to emphasize that the ECPC matrix was

---

[5] Administrative records are excluded.

[6] See footnote 2.

constructed quite independently of the heterogeneity index. The matrix therefore offers a backdrop against which to evaluate the heterogeneity index. The balance of this section analyzes the correspondence between the heterogeneity index as a quantitative indicator of production-oriented classification and the ECPC's qualitative judgment of the existing U.S. industrial classification.

A clear hypothesis emerges immediately from the structure of the ECPC matrix and the definition of the heterogeneity index. As the legend to the table indicates, a "D" in the "process" column suggests that a unique, well-defined process defines the industry. Similarly, a "D" in the "material" column indicates that the defining characteristic of the industry is a unique, homogeneous material or mix of materials used across establishments in the industry. Put simply, by assigning a "D" to an industry's process or material columns,[7] the ECPC is effectively concluding that the establishments within that industry have very similar production functions.

The heterogeneity index derived in the preceding section is similarly sensitive to the degree of homogeneity in the production functions found among an industry's establishments. In particular, the index for an industry approaches zero as the production functions of the member establishments become increasingly homogeneous.

Therefore, assuming the judgments incorporated into the ECPC matrix analysis are correct, one would expect those industries with "D" in any supply-based column to have corresponding H values that are low relative to other industries. That turns out to be the case. Among the 175 manufacturing industries in the matrix, 40 industries have a "D" reported in the "Process" and/or "Material" columns under the heading "Supply-based." Among these 40 industries, 34 have H values below the median (i.e., below 50), confirming in all but 6 cases a strict correspondence between the ECPC analysis and the heterogeneity index. Moreover, 23 of these 40 industries have indexes with values below 20 and 14 of them have index values less than 10. The heterogeneity index appears to capture quantitatively the essence of the ECPC's qualitative analysis.[8]

---

[7] It turns out that there are no industries with a "D" in the "Labor" column of the matrix.

[8] Among the 175 industries in the matrix having index values, there are 14 industries that have a "D" reported in both demand- and supply-based columns in the matrix. These "ideal" industries are well-defined by either supply or demand characteristics. Among these 14, 12 industries have H values below 50. Five have index values below 10.

A second hypothesis, symmetric with the first, can also be evaluated. One would expect that industries with high values of H should not be identified in the ECPC matrix as uniquely defined supply-based industries. In short, high values of H should not map into industries with "D" in any supply-based column in the matrix. This, too, turns out to be the case.

Heterogeneity index values above 90 are reported for 12 industries. For 11 of 12, the ECPC team left blanks in all the supply-based columns, indicating the team's judgment that these 11 industries were not supply based. More importantly, and consistent with the model of the heterogeneity index derived above, only 1 of the 12 industries has a "D" displayed in any supply-based column, steel pipe and tubes (SIC 3317). The high H for this industry may be explained by the multiple production processes indicated in the ECPC matrix. Put simply, the ECPC team and the heterogeneity index are in near unanimous agreement that these 12 outliers have little or no supply-based concept defining their boundaries.

Extending the analysis to the 38 industries having heterogeneity index values above 70 leads to precisely the same inference. Among the 38, only 3 industries were identified by the ECPC team as being defined or partially defined by supply-based criteria, that is, by the symbols P, M, or D; and among these 3, only 1, the steel pipe and tubes (SIC 3317) case noted above, has a "D" in a "supply-based" column. The other two industries are more weakly defined by supply-based criteria. The correspondence between the ECPC analysis and the heterogeneity index is quite strong.

Much more analysis needs to be conducted on the quantitative significance of the heterogeneity index but work to date suggests that inferences gleaned from the index are consistent with the ECPC's independent analysis of the basis for industry classification. In fact, given the structure of the index and the production-oriented criteria adopted by the ECPC in developing its matrix, it can be argued that the heterogeneity index formalizes in a quantitative way the production-oriented criteria adopted by the ECPC for industry classification.

It is also important again to emphasize that the ECPC matrix and the heterogeneity index were generated quite independently. ECPC team members responsible for constructing the industry matrix did not have access to the heterogeneity index results when assigning industries to the various columns in the matrix.

The evidence suggests that the heterogeneity index generates meaningful results. As a quantitative measure, it has the advantages of being simple and objective. The index holds promise as a useful diagnostic tool to support the current multinational effort to move North American industry classifications to a consistent production-oriented standard.

3. Applications to Classification Issues

There are a number of ways the heterogeneity index developed in this paper can be used to develop and maintain a production-oriented industry classification system. Some principal applications are discussed below.

(i) Given the multinational mandate to move toward a production-based set of industry accounts, the index H could be calculated for each industry as currently defined in each nation's industrial classification system. Those industries found to have either high values of H relative to other industries in the same country or rapidly rising values of H over time become prime candidates for classification review. The relative magnitude of the indexes across industries can be used to help prioritize reclassification efforts.

A caveat, however, is in order. While high values of the index indicate heterogeneity among the establishments within an industry, low values do not necessarily indicate homogeneity. It is possible that a set of establishments may have nearly identical input shares for those 10 aggregate input classes examined in this report but the detailed input types underlying the aggregates may nevertheless be quite distinct. Though expanding the set of input classes for use in the index's calculation mitigates this problem, the index is best viewed as a strong test of heterogeneity and a weak test of homogeneity. This property, however, does not in any way compromise the index's ability to identify and prioritize industries as candidates for revision; rather, it only says that the information one can obtain from the index depends on the quality, detail, and comprehensiveness of data on inputs that are available for use in the index.

(ii) In those industries with high index values, some establishment(s) may have been misassigned to the industry. If so, the misclassified establishment(s) can be identified through a straightforward application of equation (4). The heterogeneity of each establishment (H_i) from all other establishments within the industry can be calculated. Those establishments with relatively large H_i become prime candidates for review and possible industry reclassification. Recalling that the contribution (s_iH_i) of any establishment's heterogeneity to industry H is a function of its share, s_i, in industry sales, initial attention should focus on the industry's largest establishments.

(iii) In those cases where no individual establishments surface as the principal cause of high measured heterogeneity within an industry, competing proposals to separate the industry into more homogeneous subgroups can be evaluated through a rewritten form of equation (3). Assume, for example, a proposal suggests splitting an industry into v distinct establishment subgroups. The index can be used to quantify the benefits of the proposed industry division--that is, how much reduction in industry heterogeneity would result from the proposed split. The index H can be decomposed into "within subgroup" (H_w) and "among subgroup" (H_a) components:

$$(5)\quad H = \sum_{m=1}^{v} s_m \left(\sum_{i\in m} \sum_{k\in m} s_i s_k \frac{\left(\sum_j |w_{ij} - w_{kj}|\right)}{2} \right)$$

$$+ \sum_{m=1}^{v} \sum_{n=1}^{v} s_m s_n \frac{\left(\sum_j |w_{mj} - w_{nj}|\right)}{2}$$

$$= H_w + H_a ,$$

where v represents the number of distinct subgroups and w_{mj} and w_{nj} are the mean cost shares of the jth input in the mth and nth establishment groups, respectively.[9]

The H_w and H_a decomposition provides an arms-length guide to the costs and benefits of any proposed revision. The ratio H_a/H identifies the percent of industry-wide establishment heterogeneity that could be eliminated by a restructuring of industry boundaries into v groups. The proposal that leads to the highest H_a/H ratio becomes a leading candidate for implementation. Stated alternatively, since the highest H_a/H ratio corresponds to the lowest H_w/H ratio, the proposal found to have the lowest H_w/H ratio would lead to the most technologically homogeneous subgroupings.

Clearly, one can definitionally minimize heterogeneity within an entire classification system by maximizing the number of industry classes. This tautology requires no elaboration, nor

---

[9] An application to service industry data of this decomposition of the H index into H_w and H_a is reported in Frank M. Gollop, "Evaluating SIC Boundaries and Industry Change Over Time: An Index of Establishment Heterogeneity," *Proceedings*, Second Annual Research Conference, Reston, Virginia: Bureau of the Census, U.S. Department of Commerce, pp. 361-78, March 23-26, 1986.

does the point that it is not costless to expand the set of industry classes within an industrial classification system. This is precisely what gives equation (5) its operative importance. In view of an explicit or implicit restriction limiting the overall number of industrial classes, equation (5) can be used to compare the relative benefits of competing proposals to split existing industries. Stated equivalently, equation (5) can be used to minimize the overall heterogeneity within an industrial classification system, subject to a constraint on the number of desired industrial classes.

(iv) The decomposition presented in equation (5) also can be used to generate useful descriptive statistics comparing 4-, 3- and 2-digit industry aggregates. Consider, for example, the set of 4-digit subgroupings within a 3-digit industry. Equation (5) can be used to quantify how much of the 3-digit industry's measured H is due to heterogeneity within the component 4-digit industries (H_w) and how much is due to heterogeneity among the 4-digit industries (H_a). The index H_a identifies the incremental heterogeneity introduced when moving from lower to higher digit aggregates. Effectively, users can be informed about the extent of heterogeneity inherent in the use of aggregated industry data. Moreover, if one desired to form 3-digit groupings that combined 4-digit industries that were similar in terms of production processes, the index could be used to evaluate alternative 3-digit groupings.[10]

(v) The index also can support the process by which a new establishment is assigned to its appropriate industry. Assume that alternative industry assignments are proposed for a candidate establishment. Following equation (4), a value of H_i for the new establishment can be calculated with respect to each proposed industry's set of incumbent establishments. The new establishment has a technology most like those establishments in the industry for which its calculated H_i is lowest.

(vi) The index can be used as an objective yardstick to evaluate proposed industrial classifications received from the public, trade associations, or any user group. Once some experience with the index has been accumulated, those responsible for monitoring the industrial classification system may choose to adopt an upper bound threshold value for H. Proposed establishment groupings that lead to H values greater than this threshold presumptively would be unacceptable.

(vii) One particularly nice application of the index is its treatment of vertical integration. Though vertically and

---

[10] The ECPC has a report underway that discusses various principles for constructing industry "hierarchies."

nonvertically integrated establishments currently assigned to a common industry may produce identical final products, their significantly different input mixes will contribute measurably to industry H. A production-based classification system and, in particular, the application of equation (5) will differentiate vertically and nonvertically integrated establishments.

4. Enhancements and Improvements

The 175 4-digit manufacturing heterogeneity indexes constructed for this paper did quite well when evaluated against the classification standards of the ECPC matrix. This result is really quite significant given the aggregated nature of input detail used by the index. The index, recall, was constructed on a vector distinguishing only 10 input categories. Labor input was only differentiated by production versus nonproduction workers. Material input, clearly the dominant input in manufacturing, was only disaggregated among four categories: agricultural materials, mineral inputs, nondurable materials, and durable materials. The share of capital input was calculated as the simple residual of sales less payments to labor and material inputs.

The power of the index would be enhanced greatly if there were more input detail available, especially within the material and capital aggregates. Moreover, the same list of inputs was used for all industries; a more refined analysis would permit the list of inputs to vary by industrial sector. For example, if the index is to have any meaningful application to service establishments, the occupational mix within the labor aggregates needs to be identified.

5. Conclusions

Even as presently applied, however, the heterogeneity index derived in this paper can serve as a useful quantitative tool complementing the other resources available for constructing a consistent set of industrial classifications for the three North American countries. The index of heterogeneity could be used to monitor industry assignments, to reveal outlying establishments within industries, to identify rapidly changing technologies over time, to assist with industry revisions, to evaluate public proposals, and to provide users with important information regarding an industry's technological character. The heterogeneity index should find wide use as a diagnostic and descriptive statistic. The ECPC intends to make use of the heterogeneity index in one or more of the ways described in this report in work now underway on the NAICS.

Appendix for: The Heterogeneity Index: A Quantitative Tool to Support Industrial Classifcation					
Codes	D - Defines a conceptually based industry M - Multiple processes/markets P - Partial, e.g. industry V - Vertical integration is part of definition of industry **(See "Note" at the end of the Appendix)** INDEX - see text	Supply-based			
	Titles	PROCESS	MATERIAL	LABOR	INDEX
SIC	United States (1987)				
2011	Meat packing plants	DV	DV		15
2013	Sausages and other prepared meats	DV	DV		63
2015	Poultry slaughtering and processing	M	D		41
2021	Creamery butter	D	D		1
2022	Cheese, natural and processed				28
2023	Dry, condensed, evaporated dairy products				96
2024	Ice cream and frozen desserts	D			41
2026	Fluid milk	D			43
2032	Canned specialties				8
2033	Canned fruits and vegetables	P	P		57
2034	Dehydrated fruits, vegetables, soups	M			43
2035	Pickles, sauces, and salad dressings	M			34
2037	Frozen fruits and vegetables	P	P		42
2038	Frozen specialties, nec				24
2041	Flour and other grain mill products	DV	DV		15
2043	Cereal breakfast foods				0
2044	Rice milling	D	D		15
2045	Prepared flour mixes and doughs	DV	DV		57
2046	Wet corn milling	M			6
2047	Dog and cat food				48
2048	Prepared feeds, nec				98
2051	Bread, cake, and related products	P			18
2052	Cookies and crackers	P			21
2053	Frozen bakery products, except bread	P			35
2061	Raw cane sugar	DV	DV		12
2062	Cane sugar refining	DV	DV		3
2063	Beet sugar	D	D		4
2064	Candy + other confectionery products				18
2066	Chocolate and cocoa products	M	M		8
2067	Chewing gum				1
2068	Salted and roasted nuts and seeds	D			10
2074	Cottonseed oil mills	P	D		14
2075	Soybean oil mills	P	D		2
2076	Vegetable oil mills, nec	P	D		7
2077	Animal and marine fats and oils	M	M		19
2079	Edible fats and oils, nec	M			28
2082	Malt beverages	D			2
2083	Malt	D			1
2084	Wines, brandy, and brandy spirits	M	M		41
2085	Distilled and blended liquors	M			26
2086	Bottled and canned soft drinks				96
2087	Flavoring extracts and syrups, nec				22
2091	Canned and cured fish and seafoods	M	P		72
2092	Fresh or frozen prepared fish	M	P		74
2095	Roasted coffee	M	M		10
2096	Potato chips and similar snacks	P			8
2097	Manufactured ice	D	D		11

Appendix for:
The Heterogeneity Index: A Quantitative Tool to Support Industrial Classifcation

Codes
- D - Defines a conceptually based industry
- M - Multiple processes/markets
- P - Partial, e.g. industry
- V - Vertical integration is part of definition of industry **(See "Note" at the end of the Appendix)**
- INDEX - see text

	Titles	Supply-based			
SIC	United States (1987)	PROCESS	MATERIAL	LABOR	INDEX
2098	Macaroni and spaghetti	D			5
2099	Food preparations, nec				94
2311	Men's and boys' suits and coats	P	P		65
2321	Men's and boys' shirts	P	P		73
2322	Men's + boys' underwear + nightwear	P	P		39
2323	Men's and boys' neckwear	P	P		55
2325	Men's and boys' trousers and slacks	P	P		40
2326	Men's and boys' work clothing	P	P		85
2329	Men's and boys' clothing, nec				89
2331	Women's + misses' blouses + shirts	P	P		82
2335	Women's, junior's, + misses' dresses	P	P		86
2337	Women's and misses' suits and coats	P	P		60
2339	Women's and misses' outerwear, nec				76
2341	Women's and children's underwear	P	P		74
2342	Bras, girdles, and allied garments	P	P		19
2353	Hats, caps, and millinery		P		3
2361	Girls' + children's dresses, blouses	P	P		71
2369	Girls' and children's outerwear, nec				80
2371	Fur goods	D	D		0
2381	Fabric dress and work gloves				75
2384	Robes and dressing gowns	P	P		81
2385	Waterproof outerwear	M			53
2386	Leather and sheep-lined clothing	P	P		59
2387	Apparel belts				63
2389	Apparel and accessories, nec				58
2391	Curtains and draperies	P	P		84
2392	Housefurnishings				92
2393	Textile bags				95
2394	Canvas and related products		p		67
2395	Pleating and stitching				74
2396	Automotive and apparel trimmings				94
2397	Schiffli machine embroideries				64
2399	Fabricated textile products, nec				94
3312	Blast furnaces and steel mills	M			76
3313	Electrometallurgical products	D	D		89
3315	Steel wire and related products	MV	MV		66
3316	Cold finishing of steel shapes	MV	MV		61
3317	Steel pipe and tubes	MV	DV		92
3321	Gray and ductile iron foundries	D	D		33
3322	Malleable iron foundries	D	D		3
3324	Steel investment foundries	D	P		6
3325	Steel foundries	D	P		36
3331	Primary copper	M	M		94
3334	Primary aluminum	M	M		15
3339	Primary nonferrous metals, nec	M	M		100
3341	Secondary nonferrous metals				44
3351	Copper rolling and drawing	M	D		44

Appendix for:
The Heterogeneity Index: A Quantitative Tool to Support Industrial Classifcation

Codes D - Defines a conceptually based industry
M - Multiple processes/markets
P - Partial, e.g. industry
V - Vertical integration is part of definition of industry
(See "Note" at the end of the Appendix)
INDEX - see text

SIC	Titles United States (1987)	Supply-based PROCESS	MATERIAL	LABOR	INDEX
3353	Aluminum sheet, plate, and foil	M	D		4
3354	Aluminum extruded products	D	D		49
3355	Aluminum rolling and drawing	M	D		20
3356	Nonferrous rolling and drawing, nec	M			69
3357	Nonferrous wiredrawing + insulating	M			93
3363	Aluminum die-castings	P	D		25
3364	Nonferrous die-casting exc. aluminum	P	M		46
3365	Aluminum foundries	P	D		35
3366	Copper foundries	P	D		47
3369	Nonferrous foundries, nec	P	M		13
3398	Metal heat treating	D			7
3399	Primary metal products, nec				5
3511	Turbines and turbine generator sets	M			67
3519	Internal combustion engines, nec	P			36
3523	Farm machinery and equipment				30
3524	Lawn and garden equipment				51
3531	Construction machinery				42
3532	Mining machinery				43
3533	Oil and gas field machinery				81
3534	Elevators and moving stairways				62
3535	Conveyors and conveying equipment				56
3536	Hoists, cranes, and monorails				38
3537	Industrial trucks and tractors				86
3541	Machine tools, metal cutting types	M			66
3542	Machine tools, metal forming types	M			69
3543	Industrial patterns	D			20
3544	Special dies, tools, jigs + fixtures	M		M	30
3545	Machine tool accessories				40
3546	Power-driven handtools				46
3547	Rolling mill machinery				37
3548	Welding apparatus	M			47
3549	Metalworking machinery, nec				39
3552	Textile machinery				74
3553	Woodworking machinery				54
3554	Paper industries machinery				38
3555	Printing trades machinery				79
3556	Food products machinery				51
3559	Special industry machinery, nec				68
3561	Pumps and pumping equipment				49
3562	Ball and roller bearings	D			17
3563	Air and gas compressors				86
3564	Blowers and fans				55
3565	Packaging machinery				40
3566	Speed changers, drives, and gears				27
3567	Industrial furnaces and ovens	P			67
3568	Power transmission equipment				19
3569	General industrial machinery, nec				56

Appendix for:
The Heterogeneity Index: A Quantitative Tool to Support Industrial Classifcation

Codes D - Defines a conceptually based industry
M - Multiple processes/markets
P - Partial, e.g. industry
V - Vertical integration is part of definition of industry
(See "Note" at the end of the Appendix)
INDEX - see text

	Titles	Supply-based			
SIC	United States (1987)	PROCESS	MATERIAL	LABOR	INDEX
3571	Electronic computers	D			58
3572	Computer storage devices	M			67
3575	Computer terminals				67
3577	Computer peripheral equipment				60
3578	Calculating and accounting equipment				70
3579	Office machines, nec				31
3581	Automatic vending machines	M			29
3582	Commercial laundry equipment	PM			26
3585	Refrigeration and heating equipment				29
3586	Measuring and dispensing pumps	M			32
3589	Service industry machinery, nec				60
3592	Carburetors, pistons, rings, valves				23
3593	Fluid power cylinders + actuators				34
3594	Fluid power pumps and motors				23
3596	Scales and balances, exc. laboratory				54
3599	Industrial machinery, nec				21
3711	Motor vehicles and car bodies	M			14
3713	Truck and bus bodies	MV			62
3714	Motor vehicle parts and accessories				81
3715	Truck trailers	D			58
3716	Motor homes	PV			20
3721	Aircraft	M			18
3724	Aircraft engines and engine parts				21
3728	Aircraft parts and equipment, nec				24
3731	Ship building and repairing	M			33
3732	Boat building and repairing				80
3743	Railroad equipment				27
3751	Motorcycles, bicycles, and parts				49
3761	Guided missiles and space vehicles	D			38
3764	Space propulsion units and parts				17
3769	Space vehicle equipment, nec				49
3792	Travel trailers and campers	P			62
3795	Tanks and tank components				4
3799	Transportation equipment, nec				94

Note: Entries in the first three columns under the "supply-based" heading are taken from Appendix B in ECPC Report No. 1, "Economic Concepts Incorporated in the Standard Industrial Classification Industries of the United States." The entries are defined, and the process for preparing them is described in ECPC Report No. 1.