Appendix: The NBER Immigration, Trade, and Labor Markets Data Files

John M. Abowd

The National Bureau of Economic Research Immigration, Trade, and Labor Markets Data Files were developed from public data sources to facilitate industry-based research on the effects of international trade and immigration on labor markets in the United States. Many of the papers in this volume make use of data derived from these files.[1] The purpose of this paper is to document the sources and methods used to prepare the data files.

There are three basic types of files. Trade data files contain information organized on an industry basis for U.S. manufacturing industries. Industry immigration files contain information on work force characteristics, including immigrants, organized on an industry basis for all U.S. industries. Area immigration files contain information on work force characteristics, including immigrants, organized on a state and standard metropolitan statistical area (SMSA) basis.

The Trade Data File contains shipments, a shipments deflator, value added, employment, payroll, real capital stock, imports, exports, unionization, and immigrant ratios for 450 four-digit (1972 Standard Industrial Classification) industries. The files provide annual data covering the period from 1958 to 1986. The Industry Immigration File contains information on the education, demographics, and immigrant proportions in the labor force of 292 Census Industrial Classification (CIC) industries. The area immigrant files contain information on immigrant proportions in the fifty states (State Immigrant File) and SMSAs (SMSA Immigrant File).

Overall Organization of the Data

Table A.1 contains a list of the variable names, short definitions, and basic sources for the variables in the Trade Data File. Table A.2 contains a comparable list for the variables in the Industry Immigration Data File. Table A.3 contains a list for the variables in the Area Immigration Data file. Variable names shown in tables A.1, A.2 and A.3 are used throughout this paper to refer to the variables in the files.

Annual Survey of Manufactures Data

The industry level production, factor use, and price index data were prepared by Wayne B. Gray as a part of the NBER's Productivity Project. The industry definitions conform to the 1972 Standard Industrial Classification. These definitions are used throughout the Trade and Immigration Data Files.

The initial version of the industry level production, factor use, and price index data, covering the period from 1958 to 1976, was developed as a joint

Table A.1   Names, Short Definitions, and Sources for Variables in the NBER Trade Data File

Acronym	Short Definition	Units	Source
SIC	Four-digit industry identifier		1972 SIC[a]
YEAR	Year[b]		
VSHIPPED	Value of product shipments	$millions	ASM[c]
SHIPDEFL	Shipments deflator	1972 = 1.00	BLS[d]
VADDED	Value added by manufacture	$millions	ASM
ALL_WRK	All employees, number	thousands	ASM
PROD_WRK	Production workers, number	thousands	ASM
PROD_HRS	Production workers, hours	millions	ASM
ALL_PAY	All employees, payroll	$millions	ASM
PROD_WAG	Production workers, payroll	$millions	ASM
ALL_UN	Percentage of all employees unionized	%	CPS[e]
PROD_UN	Percentage of production workers unionized	%	CPS
RCAPSTCK	Real capital stock in 1972 dollars	$millions	SRI-Penn[f]
IMPORTS	Customs value of imports	$millions	TMS[g]
EXPORTS	Free alongside ship value of exports	$millions	TMS
IMMRAT	Ratio of immigrants to labor force in SIC ratio		NBER Imm.[h]

[a] Standard Industrial Classification, manufacturing only.
[b] The year includes the century (e.g., 1971).
[c] *Annual Survey of Manufactures Statistics for Industry Groups and Industries.*
[d] Bureau of Labor Statistics (unpublished).
[e] Current Population Survey, May public use data.
[f] University of Pennsylvania, Bureau of the Census, SRI, Incorporated, 1958–76. Bureau of Industrial Economics, Commerce Department, 1977–86.
[g] Bureau of Labor Statistics Trade Monitoring System.
[h] NBER Immigration Data Files, see table A.2.

Table A.2 **Names, Short Definitions, and Sources for Variables in the NBER Industry Immigration Data Files**

Acronym	Short Definition	Units	Source
SIC	Four-digit industry identifier		1972 SIC[a]
IMRAT60	Immigrants/total employment 1960	Proportion	CP60 1/100[b]
IMRAT70	Immigrants/total employment 1970	Proportion	CP70 1/100[c]
ARAT70	Asian immigrants/total employment 1970	Proportion	CP70 1/100
BRAT70	Black immigrants/total employment 1970	Proportion	CP 70 1/100
WRAT70	White immigrants/total employment 1970	Proportion	CP70 1/100
MRAT70	Mexican immigrants/total employment 1970	Proportion	CP70 1/100
ORAT70	Other hispanic immigrants/total employment 1970	Proportion	CP70 1/100
RRAT70	Recent (last 2 years) immigrants/total employment 1970	Proportion	CP70 1/100
IMRAT80	Immigrants/total employment 1980	Proportion	CP80 A[d]
ARAT80	Asian immigrants/total employment 1980	Proportion	CP80 A
BRAT80	Black immigrants/total employment 1980	Proportion	CP80 A
WRAT80	White immigrants/total employment 1980	Proportion	CP80 A
MRAT80	Mexican immigrants/total employment 1980	Proportion	CP80 A
ORAT80	Other Hispanic immigrants/total employment 1980	Proportion	CP80 A
RRAT80	Recent (last 2 years) immigrants/total employment 1980	Proportion	CP80 A
CIC	Census Industrial Code		1970 or 1980
YEAR	Year of Census		
PR1624	Labor force age 16–24	Proportion	CP80 A
PR2COLL	Labor force with at least 2 years college	Proportion	CP80 A
PRFEMALE	Females in labor force	Proportion	CP80 A
PRBLACK	Blacks in labor force	Proportion	CP80 A
PROD_UN	Production worker unionization rate	Proportion	See table A.1

[a] Standard Industrial Classification, manufacturing only.
[b] Census of Population 1960 1/100 Public Use Sample.
[c] Census of Population 1970 1/100 Public Use County Group Sample.
[d] Census of Population 1980 extracts from the A Sample.

project by the University of Pennsylvania, the Bureau of the Census, and SRI, Incorporated (called SRI-Penn hereafter). The SRI-Penn data are documented in Andrews and Zabala (1984) and the references therein.[2] SRI-Penn variables from the *Annual Survey of Manufactures (ASM) Statistics for Industry Groups and Industries* have not been changed. The SRI-Penn data were extended to

Table A.3   **Names, Short Definitions, and Sources for Variables in the NBER Area Immigration Data Files**

Acronym	Short Definition	Units	Source
STATE	Census state code		CP70/CP80[a]
SMSA	Census metropolitan area codes[b]		CP70/CP80
IMRAT70	Immigrants/total employment 1970	Proportion	CP70 1/100[c]
IMRAT80	Immigrants/total employment 1980	Proportion	CP80 A[d]

[a] Census of Population 1970 1/100 Public Use County Group Sample or Census of Population 1980 extracts from the A Sample.
[b] Codes for metropolitan areas vary between 1970 and 1980; see the Census of Population documentation cited in the references for appropriate codes.
[c] Census of Population 1970 1/100 Public Use County Group Sample.
[d] Census of Population 1980 extracts from the A Sample.

1986 using *ASM Statistics for Industry Groups and Industries*, table 2 (1986 and earlier years).

The implementation of consistent bridges to recode all industry data to a 1972 SIC basis was a major problem in the creation of the four-digit industry-based data. The SRI-Penn data (1958–76) were based on 1967 SIC definitions from the defunct *Industry Profiles* data base (U.S. Department of Commerce 1971, 1978). The SRI-Penn project recoded the 1967 SIC industry data to a 1972 SIC basis using the many-to-many bridge based on the industry translation table published as a part of the Census of Manufactures (U.S. Department of Commerce 1976). This bridge assigned to the 1972 SIC the proportion of the 1967 SIC-based variable that was appropriate based on the proportion of 1972 value added represented by the 1967 SIC-based industry.

The data from 1977 to 1986 were restated on a 1972 basis as follows. Four-digit 1977 SIC-based data were recoded to a 1972 basis using the many-to-many bridge that accompanied the 1977 Census of Manufactures (U.S. Department of Commerce 1981). This bridge assigns to the 1972 SIC the proportion of the 1977 SIC that was appropriate based on the proportion of 1972 value added represented by the 1977 SIC-based industry. (Although there are only 448 1977-based four-digit SIC industries, the NBER files continue to use the 450 1972-based four-digit SIC industries.)

Some missing data were imputed because the values for a particular four-digit SIC were suppressed from the original published ASM tables in order to avoid the breach of establishment confidentiality. For the period from 1958 to 1976, the SRI-Penn data estimated the value of missing data by a combination of remainder assignment and interpolation between the Census of Manufactures years (see Andrews and Zabala 1984, 7).

For the period from 1977 to 1986 in the NBER Trade Data File, the following procedure was used. The complete two-digit industry group data and the available three-digit industry group data were used to impute the missing three-digit data by calculating the remainder of the two-digit classification after subtracting all available three-digit data. The two-digit remainder was

allocated to missing three-digit industry groups by imputation based on available three-digit data in surrounding years. The complete three-digit industry group data (including imputations) and the available four-digit industry data were used to impute missing four-digit data following the same remainder and allocation process as used for the missing three-digit data.

The detailed definitions of ASM variables can be found in Appendix A-1 of *ASM Statistics for Industry Groups and Industries* (1986 and earlier years). The ASM is a survey of manufacturing establishments with a sampling frame derived from the Census of Manufactures conducted every five years. The following variables were used directly from ASM Table 2. VSHIPPED is the total annual value of industry shipments in millions of current dollars. VADDED is value added by manufacture in millions of current dollars. ALL_WRK is the annual average employment of full- or part-time persons in thousands. PROD_WRK is the annual average full- and part-time employment of production workers (through the line supervisor level) in thousands. PROD_HRS is the total annual hours of production workers in millions. ALL_PAY is the total annual value of gross earnings paid to all employees in millions of current dollars (includes salaries, wages, commissions, bonuses, vacation pay, sick pay, and compensation in kind as used in calculating federal withholding taxes). PROD_WAG is the total annual value of gross earnings paid to production workers in millions of current dollars.

The estimated real value of the capital stock (in 1972 dollars) is based on the SRI-Penn data for the period 1958–76. The methods used to develop this series are described in Andrews and Zabala (1984, 10–15). The variable RCAPSTCK for the period 1958–76 is an accumulation of the depreciated, constant dollar plant and equipment investment series for the industry.

For the period from 1977 to 1986, the real capital stock was estimated using unpublished information from the Industry Capital Stocks data base, which provides deflators for new investment and separate measures of real plant and equipment capital stocks for three-digit industries.[3] These data were used to estimate implicit depreciation rates for plant and equipment investments in the three-digit SIC industries. The price deflators and depreciation rates for each three-digit industry were assigned to all four-digit industries within that SIC. The published ASM new capital expenditures were divided into plant and equipment using unpublished data from the Industry Capital Stocks data base. The four-digit SIC real plant and equipment stocks were estimated separately using the formula

real plant stock = (1 − depreciation rate) (previous year plant stock)
+ (previous year new plant spending)/
(previous year plant deflator),

and similarly for the real equipment stock. The variable RCAPSTCK for the period from 1977 to 1986 is the sum of the real plant and real equipment stocks.

Because the Industry Capital Stocks data base ends in 1982, deflators for

new plant and equipment were derived from the Detailed Investment by Industry data base. These data are available only at the two-digit level. The four-digit industry was assigned the new plant and equipment deflator from its two-digit industry group for 1983 forward.

The shipments deflator for the period 1958–76 comes from the SRI-Penn data. For the period from 1977 to 1986, the deflator is based on unpublished data from the Bureau of Economic Analysis (BEA) in the Department of Commerce. The BEA four-digit SIC shipment deflators are based on the Bureau of Labor Statistics product price indices. The BLS product price indices are based on the seven-digit detailed product class codes, which are aggregated to a four-digit SIC basis. There have been many changes in the product definitions and available price indices throughout the years. There are apparently no comprehensive descriptions of these series within the BLS or the BEA. In recent years (varying depending on the industry), the BLS has created industry price indices (as opposed to product price indices) at the four-digit SIC level. The BLS industry price indices were used whenever possible. The shipments deflator is the variable SHIPDEFL.[4]

International Trade Data

Since most domestic U.S. data are maintained on an industrial classification basis, primarily the SIC, the NBER Trade Data File was designed to provide import and export data on an industrial basis. The difficulty with this approach is that the SIC system requires information about the product type and the manufacturing method while the basic classification system used for imports and exports identifies only the product type. As a consequence, the most detailed SIC-based classifications of imports and exports group together products that span several four-digit SIC categories. The noncomparability of SIC-based import, export, and production data, as published, required the development of alternative industry-based estimates of international trade values.

The annual industry measures of import and export value that appear in the NBER Trade Data File were derived from the Bureau of Labor Statistics *Trade Monitoring System* (1979, 1983a, 1983b), the Bureau of the Census *U.S. Commodity Exports and Imports as Related to Output* (1965/64 and 1972/71), the ASM data described above, and the *ASM Value of Product Shipments* (1982–85).

For the period from 1972 to 1981, the Bureau of Labor Statistics maintained a collection of time series called the *Trade Monitoring System* (TMS) that provided annual information on U.S. imports, the ratio of imports to new supply, exports, and the ratio of exports to domestic shipments on a modified four-digit SIC-based industry classification. Schoepfle (1982) reports the development and uses of the TMS data. His appendix contains numerous details of the calculations. Bennett (1982, available on request from the Department of Labor) reports the details of the TMS data base construction.

The TMS was based on a very careful attempt to construct domestic shipment data that were as comparable as possible to the most detailed industrial classification–based import and export data. The NBER Trade Data File follows the TMS methodology to extend the series backward to 1958 and forward to 1985. The TMS methods were approximated for the period from 1958 to 1972, using 1972 as a splice year. The TMS methods were used exactly for 1982–85.

For imports, the TMS defined 317 mutually exclusive manufacturing groups in a classification system called the MSIC (for import SIC). The MSIC covers all four-digit 1972 SIC-based manufacturing industries that can be distinguished from the automatic application of the concordance relating U.S. Department of Commerce Tariff Schedules of the United States, Annotated (TSUSA), to Product Class code (five-digit 1977-based SIC), which forms the basis for the tables in the publication *U.S. Commodity Exports and Imports as Related to Output* and the Department of Commerce online data base of official international trade statistics.[5] The basic import data are collected continuously at the port of entry to the United States by the Bureau of the Census and are classified according to the TSUSA as a part of the Customs process. The concordance between TSUSA and Product Class codes is updated monthly as TSUSA codes are created or eliminated. Annual import data aggregated to a four-digit SIC basis are published regularly by the Census Bureau on the basis of the current TSUSA–to–Product Class concordance and summarized biennially in the publication *U.S. Commodity Exports and Imports as Related to Output*. In this publication, a modified four-digit SIC, similar to the MSIC, is used to distinguish the industries.[6]

For exports, the TMS defined 370 mutually exclusive four-digit SIC manufacturing industries in a classification system called the XSIC (for export SIC). The classification is based on all 1972-based SIC industries that can be distinguished from an automatic application of the U.S. Department of Commerce concordance between Schedule B commodity numbers and five-digit SIC-based Product Class codes. The basic export data are collected at the port of departure from the United States by the Bureau of the Census and are classified according to the Schedule B number as a part of the coding process for international trade statistics. The concordance between Schedule B number and Product Class codes is updated monthly as Schedule B numbers are created or eliminated. Annual export data aggregated to a four-digit SIC basis are published regularly by the Census Bureau on the basis of the current Schedule B–to–Product Class concordance and summarized biennially in the publication *U.S. Commodity Exports and Imports as Related to Output*. In this publication, a modified four-digit SIC, similar to the XSIC, is used to distinguish the industries.[7]

The import data in the NBER Trade Data File were assembled for the 317 basic MSIC industries. For each such industry, imports are defined as the customs value of imports in millions of current dollars. The customs value generally excludes transportation and insurance costs from the foreign port. For

the period from 1958 to 1971, imports were reported in *U.S. Commodity Exports and Imports as Related to Output* using an old definition. For 1972 to 1981, imports were used directly from the TMS. For 1982 to 1985, imports were reported in the U.S. Department of Commerce online data base of official statistics from the table *All Items in U.S. Imports for Consumption from World* (1986b) using a consistent definition but a revised concordance. The year 1972 was used as a splice to make the 1958–71 values conform to the 1972 definitions and concordance.[8] The year 1981 was used as a splice to make the 1982–85 values conform to the 1981 concordance.[9]

The MSIC-based value of domestic shipments for the industry or industry group corresponding to the MSIC was taken from the TMS for the years 1972–81. For the years 1958–71, the MSIC-based value of domestic shipments was imputed using the following technique. The value of industry shipments for the 1972 SIC-based four-digit industries was obtained from the ASM data discussed above. Using the concordance between MSIC and SIC developed at the NBER, the ASM value of shipments data were merged with the pre-1972 MSIC-based import data. Then the SIC-based shipments data for the years 1958–71 were multiplied by the ratio of MSIC-based value of shipments in 1972 to SIC-based value of shipments in 1972. For the years 1982–85, the TMS definitions of MSIC-based value of domestic shipments, which are based on comparable five-digit Product Class codes, were constructed directly from the table 1 values in *ASM Value of Product Shipments*. Data from the five-digit value of product shipments table represent sampling estimates of the net sales value freight-on-board at the point of manufacture in millions of current dollars (excludes discounts, transportation costs, and excise taxes).

For each MSIC-based industry, the import penetration ratio (IPR) was defined as the ratio of imports to the sum of imports and (MSIC-based) domestic shipments. Using the MSIC-to-SIC concordance, the import penetration ratio was merged to the data for the 450 1972 SIC-based industries for the years 1958–85. The value of imports for the SIC-based industry was set to zero if there was no concordant MSIC. Otherwise, IMPORTS in the Trade Data File was calculated as the product of ASM shipments (the variable vSHIPPED) and the ratio IPR/(1 − IPR). In the Trade Data File, the basic import penetration ratio can be reconstructed exactly by forming the ratio IMPORTS/(vSHIPPED + IMPORTS).

Using 1972 as the reference year, the quality of the MSIC-to-SIC concordance at the four-digit level can be assessed by considering the summary statistics in table A.4. Slightly more than 50 percent of the industry value of shipments can be matched exactly. An additional 48 percent of industry shipments can be match by combining SICs, using special MSICs, and other concordance arrangements. Overall, import statistics are available for 98.3 percent of all shipments.

The export data in the NBER Trade Data File were assembled for the 370 basic XSIC industries. For each such industry, exports are defined as the free

Table A.4   **Quality of the Concordance between Import-based MSICs and Conventional 1972-based SICs**

Category	Percentage of 1972 Shipments	Percentage of 1972 Industries
Exact match	50.9	55.8
Combined 2 or more SICs into MSIC	30.7	28.7
Special code required	15.0	10.7
Other basis for concordance	1.6	1.7
Excluded from TMS	1.7	3.1

Source: NBER Trade Data File and author's calculations.

alongside ship value of exports in millions of current dollars. The free alongside ship value generally includes transportation and insurance costs from the point of manufacture to the port of departure from the United States. For the period from 1958 to 1971, the export value was reported in *U.S. Commodity Exports and Imports as Related to Output*. For 1972–81, the export value was taken directly from the TMS. For 1982–85, the export value was reported in the U.S. Department of Commerce online data base of official statistics from the table *All Items in U.S. Domestic Exports to World* (1986a) using a consistent definition but a revised concordance. The year 1972 was used as a splice to make the 1958–71 values conform to the 1972 definitions and concordance.[10] The year 1981 was used as a splice to make the 1982–85 values conform to the 1981 concordance.[11]

The XSIC-based value of domestic shipments for the industry or industry group corresponding to the XSIC was taken from the TMS for the years 1972–81. For the years 1958–71, the XSIC-based value of domestic shipments was imputed using the following technique. The value of industry shipments for the 1972 SIC-based industries was obtained for the four-digit industry from the ASM data discussed above. Using the concordance between XSIC and SIC developed at the NBER, the ASM value of shipments data were merged with the pre-1972 XSIC-based export data. Then the SIC-based shipments data for the years 1958–71 were multiplied by the ratio of XSIC-based value of shipments in 1972 to SIC-based value of shipments in 1972. For the years 1982–85, the TMS definitions of XSIC-based value of domestic shipments, which are based on comparable five-digit Product Class codes, were constructed directly from the table 1 values in *ASM Value of Product Shipments*.

For each XSIC-based industry, the export supply ratio (XS) was defined as the ratio of exports to (XSIC-based) domestic shipments. Using the XSIC-to-SIC concordance, the export supply ratio was merged to the data for the 450 1972 SIC-based industries for the years 1958–85. The value of exports for the SIC-based industry was set to zero if there was no concordant XSIC. Otherwise, EXPORTS in the Trade Data File was calculated as the product of ASM shipments (the variable VSHIPPED) and the variable XS. In the Trade Data File,

Table A.5   **Quality of the Concordance between Export-based XSICs and Conventional 1972-based SICs**

Category	Percentage of 1972 Shipments	Percentage of 1972 Industries
Exact match	61.9	74.2
Combined 2 or more SICs into XSIC	33.8	18.7
Special code required	1.7	2.0
Other basis for concordance	.7	1.1
Excluded from TMS	1.8	4.0

Source: NBER Trade Data File and author's calculations.

the basic export supply ratio can be reconstructed exactly by forming the ratio EXPORTS/VSHIPPED.

Using 1972 as the reference year, the quality of the XSIC-to-SIC concordance at the four-digit level can be assessed by considering the summary statistics in table A.5. Almost 62 percent of the industry value of shipments can be matched exactly. An additional 36 percent of industry shipments can be matched by combining SICs, using special XSICs, and other concordance arrangements. Overall, export statistics are available for 98.2 percent of all shipments.[12]

Current Population Survey Data

The industry unionization data were derived using the methods of Freeman and Medoff (1979). The Freeman-Medoff estimates were updated into the 1980s but cannot be used for the period from 1958 to 1972 because the estimates rely on the unionization questions from the May Current Population Survey (CPS).

Union membership percentages were calculated as

$$U_j = \left(\sum_i A_{ij} W_{ij} \Big/ \sum_i W_{ij} \right) 100,$$

where U_j is percentage of workers in Census industry j who are unionized; $A_{ij} = 1$ if worker i is employed and in a union and is zero otherwise; and W_{ij} is the CPS sampling weight attached to worker i. Separate unionization rates were calculated for all workers and for production workers. Production workers were defined as employed individuals in the following occupations: craftsmen and kindred, operatives except transport, transport operatives, nonfarm laborers, private household, all other service, farm laborers and foremen. The values for 1974 were based on the 1973, 1974, and 1975 May surveys. The values in the Trade Data File for 1958–73 are identical to the 1974 values. The values for 1980 were based on the 1979, 1980, and 1981 May surveys.

The values for 1984 were based on the 1984 May survey. All other years were linearly interpolated.

The unionization data for detailed CIC from the May Current Population Survey were matched to the 1972 SIC-based industry data using a one-to-many concordance between 1970 CICs and 1972 SICs. The variable ALL_UN contains the estimated overall unionization rate from the concordant CIC. The variable PROD_UN contains the estimated production unionization rate from the concordant CIC.

Census of Population and Housing Data

Public use microdata samples from the 1960, 1970, and 1980 decennial Census of Population and Housing were used to derive estimates of the ratio of immigrant employees to total employees by industry for the Census years. In 1970 and 1980, these immigrant ratios are also available by racial and ethnic groups. The 1970 and 1980 Censuses were also used to create industry and area data on labor force characteristics.

The 1960 immigrant ratio was created from the 1960 Census of Population and Housing using the 1/100 Public Use Sample (Bureau of the Census 1975). The numerator of the ratio is the count of all immigrants employed in the detailed 1960 CIC industry. The denominator is the count of all individuals employed in the detailed CIC. Using a one-to-many concordance between the 1960 CIC and the 1972 four-digit SIC (for the 450 manufacturing industries only), the variable IMMRAT60 was created on an SIC basis from the concordant CIC.

The 1970 immigrant ratios were created from the 1970 Census of Population and Housing using the 1970 5% Data County Group Sample, which is a 1/100 sample of the U.S. population (Bureau of the Census 1972). The numerator of the ratio is the count of all immigrants (or all immigrants in the particular racial/ethnic group) employed in the detailed 1970 CIC industry. The denominator is the count of all individuals employed in the detailed CIC industry. Using a one-to-many concordance between the 1970 CIC and the 1972 four-digit SIC (for the 450 manufacturing industries only), the immigrant ratios for 1970 were created on an SIC basis from the concordant CIC. The immigrant groups used were all immigrants (IMMRAT70), Asian immigrants (ARAT70), black immigrants (BRAT70), white immigrants (WRAT70), Mexican immigrants (MRAT70), other Hispanic immigrants (ORAT70), and all immigrants who arrived within the last two years (RRAT70). The Asian, black, and white racial groups are mutually exclusive and include Hispanics as appropriate. The two Hispanic ratios are mutually exclusive.

The 1980 immigrant ratios were created from the 1980 Census of Population and Housing using the A Sample, which is a 5/100 sample of the U.S. population (Bureau of the Census 1983a, 1983b). The numerator of the ratio is the count of all immigrants (or all immigrants in the particular racial/ethnic

group) employed in the detailed 1980 CIC industry. The denominator is the count of all individuals (from a 1/1,000 random subsample of the A Sample) employed in the detailed CIC industry (multiplied by fifty to reflect the different sampling rates). Using a one-to-many concordance between the 1980 CIC and the 1972 four-digit SIC (for the 450 manufacturing industries only), the immigrant ratios for 1980 were created on an SIC basis from the concordant CIC. The immigrant groups used were all immigrants (IMMRAT80), Asian immigrants (ARAT80), black immigrants (BRAT80), white immigrants (WRAT80), Mexican immigrants (MRAT80), other Hispanic immigrants (ORAT80), and all immigrants who arrived within the last two years (RRAT80).

The other variables in the Industry Immigration Data Files are labor force characteristics by industry in 1980. The variable PR1624 is the proportion of persons age 16–24 employed in the detailed CIC industry. The variable PR2COLL is the proportion of employed persons with two or more years of college in the detailed CIC industry. The variable PRFEMALE is the proportion of females employed in the detailed CIC industry. The variable PRBLACK is the proportion of blacks employed in the detailed CIC industry. There are 292 detailed CIC industries represented. For manufacturing industries, only a one-to-many concordance between the 1980 CIC and the 1972 four-digit SIC was used to assign the labor force characteristics of the CIC to its concordant SIC.

The Area Immigration Data Files contain the variables IMMRAT70 and IMMRAT80, as defined in the Industry Immigration Data Files, except that the numerator and denominator of the ratios were calculated for immigrants and all employed individuals in states and SMSAs. The states were defined using the FIPS State Code for both Census years. The SMSAs were defined by the SMSA code in the 1980 Census and by the area and subarea codes in the 1970 Census. See the technical documentation in the references for code lists.

Availability

The NBER Immigration, Trade, and Labor Markets Data Files are available on high density floppy disk (Strata™ format or ASCII format) or computer tape (SAS™ format) from the author (address requests to NBER, Labor Studies Program, 1050 Massachusetts Avenue, Cambridge, MA 02138). The files are also available on Internet from the author (contact JMA@CORNELLA.CIT.CORNELL.EDU on Internet or JMA@CORNELLA on Bitnet).

Notes

1. These data are used in the following papers in this volume: Abowd and Freeman, Abowd and Lemieux, Collins, Freeman and Katz, Kuhn and Wooton, and Leonard and McCulloch.

2. The SRI-Penn data files and documentation are available on request from Stephen Andrews at the Center for Economic Studies in the Bureau of the Census.

3. Information regarding the Industry Capital Stock data base is available from the Bureau of Industrial Economics in the U.S. Department of Commerce. The data base contains information on a three-digit SIC basis.

4. Only a subset of the annual industry data developed for the NBER Productivity Project are available in the NBER Trade Data File. For additional information about the ASM and related industry data, contact Wayne B. Gray at the National Bureau of Economic Research.

5. The basic classification systems for imports and exports changed in 1988. The Department of Commerce online data base currently produces tables of imports and exports by industry that are created using a different set of concordances than described in this paper. The tables from the Commerce Department online data base used in the NBER Trade Data File were based on the old concordances, described herein.

6. The *Trade Monitoring System* (1983b) describes the concordance procedure as follows: "For the purpose of relating imports to output, individual TSUSA commodity numbers are assigned to the five-digit SIC-based Product Class from the numerical list of manufactured products, 1977 Census of Manufactures, which contains the same products as the TSUSA number. In cases where the TSUSA numbers include items which should be classified in two or more SIC-based output codes, an assignment is made to the SIC-based output code to which the principal content of the TSUSA appears to belong, where such an assignment will not significantly overcount the classification to which the TSUSA number belongs. In cases where it appears that distortions will result from an assignment of an entire TSUSA number to a single SIC-based output code, the principal SIC-based output product classes are combined to form an SIC-based import code and TSUSA numbers are assigned to the combination."

7. The *Trade Monitoring System* (1983a) describes the concordance between Schedule B numbers and Product Class codes as follows:

> For the purpose of relating exports to output, individual Schedule B commodity numbers are assigned to the five-digit SIC-based Product Class from the numerical list of manufactured products, 1977 Census of Manufactures, which contains the same products as the Schedule B number. In cases where the Schedule B numbers include items which should be classified in two or more SIC-based output codes, an assignment is made to the SIC-base output code to which the principal content of the Schedule B number appears to belong, where such an assignment will not significantly overcount the classification to which the Schedule B number is assigned or undercount the other classification to which it partially belongs. In cases where it appears that distortions will result from an assignment of an entire Schedule B number to a single SIC-based output code, the principal SIC-based output product classes are combined to form an SIC-based export code and the schedule B numbers are assigned to the combination.

8. That is, the values from 1958 to 1971 were multiplied by the ratio of the 1972 TMS import value to the 1972 *U.S. Commodity Exports and Imports as Related to Output* value.

9. That is, the values from 1982 to 1985 were multiplied by the ratio of the 1981 TMS import value to the 1981 U.S. Department of Commerce official statistic.

10. That is, the values from 1958 to 1971 were multiplied by the ratio of the 1972 TMS export value to the 1972 *U.S. Commodity Exports and Imports as Related to Output* value.

11. That is, the values from 1982 to 1985 were multiplied by the ratio of the 1981 TMS export value to the 1981 U.S. Department of Commerce official statistic.

12. The international trade data were constructed by the author. For further information, contact the author at the National Bureau of Economic Research.

References

Andrews, Stephen, and Craig Zabala. 1984. *Documentation of the SRI-Penn Manufacturing Industry Dataset Developed by David L. Crawford, Gary Fromm, Lawrence R. Klein, and Frank C. Ripley.* Technical Notes, Center for Economic Studies. Washington, D.C.: Bureau of the Census, January.
Bennett, Norman. 1982. *Trade Monitoring System, Technical Note, Import Penetration and Export Proportion Data Bases.* Washington, D.C.: Bureau of Labor Statistics, Division of Foreign Labor Statistics and Trade, November.
Freeman, Richard B., and James L. Medoff. 1979. New Estimates of Private Sector Unionism in the U.S. *Industrial and Labor Relations Review* 32, no. 2 (January):143–74.
Gray, Wayne. 1986. *Productivity vs. OSHA and EPA regulation.* Ann Arbor, Mich.: UMI Research Press.
———. 1987. The Cost of Regulation: OSHA, EPA and the Productivity Slowdown. *American Economic Review* 77, no. 5 (December):998–1006.
Schoepfle, Gregory. 1982. Imports and Domestic Employment: Identifying Affected Industries. *Monthly Labor Review* 105, no. 8 (August): 13–26.
U.S. Department of Commerce. 1986a. *All Items in U.S. Domestic Exports to World, 1981–85* [online data base of official statistics]. Washington, D.C.: U.S. Department of Commerce.
———. 1986b. *All items in U.S. Imports for Consumption from World, 1981–85* [online data base of official statistics]. Washington, D.C.: U.S. Department of Commerce.
U.S. Department of Commerce. Bureau of the Census. 1971. *Industry Profiles, 1958–1969.* Washington, D.C.: U.S. Government Printing Office, October.
———. 1972. *Public Use Samples of Basic Records from the 1970 Census: Description and Technical Documentation.* ICPSR Study no. 0018. Washington, D.C.: Bureau of the Census.
———. 1975. *Public Use Sample of Basic Records from the 1960 Census: Description and Technical Documentation.* ICPSR Study no. 7756. Washington, D.C.: Bureau of the Census, April.
———. 1976. *1972 Census of Manufactures.* Washington, D.C.: U.S. Government Printing Office.
———. 1978. *Industry Profiles, Annual Survey of Manufactures.* Washington, D.C.: U.S. Government Printing Office, June.
———. 1973, 1974, 1975, 1979, 1984. *Current Population Survey May 1979 Technical Documentation.* ICPSR Study nos. 7936, 7937, 7938, 7974, 8461. Washington, D.C.: Bureau of the Census.
———. 1981. *1977 Census of Manufactures.* Washington, D.C.: U.S. Government Printing Office, August.
———. 1983a. *Census of Population and Housing, 1980: Public-Use Microdata Sample (A Sample)* [machine-readable data file]. ICPSR Study no. 8101. Washington, D.C.: Bureau of the Census.
———. 1983b. *Census of Population and Housing, 1980: Public-Use Microdata Samples Technical Documentation.* Washington, D.C.: Bureau of the Census, Data User Services Division.

————. Annual. *Annual Survey of Manufactures Statistics for Industry Groups and Industries.* Washington, D.C.: U.S. Government Printing Office.

————. Annual. *Annual Survey of Manufactures Value of Product Shipments.* Washington, D.C.: U.S. Government Printing Office.

————. Biennial. *U.S. Commodity Exports and Imports as Related to Output.* Washington, D.C.: U.S. Government Printing Office.

U.S. Department of Commerce. Bureau of Economic Analysis. Monthly. *Survey of Current Business.*

U.S. Department of Labor. Bureau of Labor Statistics. Office of Productivity and Technology. Division of Foreign Labor Statistics and Trade. 1979. *Trade Monitoring System Import Penetration by Four-Digit SIC-based Manufacturing Commodity Group* [unpublished computer listing]. March.

————. 1983a. *Trade Monitoring System U.S. Exports and Related Output by Four-Digit SIC-based Commodity Group Manufactures Exports, Product Shipments, and Exports to Shipments Ratio, 1972–1981* [unpublished computer listing]. November.

————. 1983b. *Trade Monitoring System U.S. Imports and Related Output by Four-Digit SIC-based Commodity Group Manufactures Imports, Product Shipments, New Supply, and Import Penetration, 1972–1981* [unpublished computer listing]. November.