International Trade and California's Economy:

Summary of the Data

by

Professor Dwight M. Jaffee

Fisher Center for Real Estate and Urban Economics
Haas School of Business
University of California
Berkeley CA 94720-1900

Tel: 510-643-6105

Email: jaffee@haas.berkeley.edu

Acknowledgment

International Trade and California's Economy:
Summary of the Data

Abstract

This paper presents most of the primary data sets that have been assembled for use in the project, *Foreign Trade and California's Growth.* The data cover the employment, sales, and international trade (imports and exports) activities of all industries for both California and the US as a whole. The data are available primarily at the 3-digit SIC code level, although some 4-digit SIC code data are also used. Historical data are available for at least five previous years.

The discussion summarizes the key features of the California economy that are illustrated by the data and describes the statistical algorithms that were developed and applied in order to create a complete and consistent data set. These data are used to carry out statistical tests of various hypotheses in other papers of the project.

1   Introduction

A large data set is needed to determine the effects of international trade on the California economy. The data should cover at least the employment, sales, and international trade (imports and exports) of all industries for both California and the US as a whole. The data should be available at a detailed industry level, preferably to the level of 4-digit SIC codes. The industry definitions should be consistent across the employment, sales, and trade data. Historical data should be available for five years or more.

This paper presents most of the main data sets we have assembled for this purpose. The discussion summarizes the key features of the California economy that are illustrated by the data and describes the statistical algorithms that were developed and applied in order to create a complete and consistent data set. These data are used in other papers associated with this project to carry out statistical tests of various hypotheses.

2   Employment Data for California and the US

To understand the links between international trade and the state economy, we first need accurate data on employment, sales, and trade flows. This is not as simple as it might first appear. There are issues regarding the timeliness, consistency, and completeness of the data even at the US level and especially for California (or other states). In the following discussion, we use the computer industry as an example to illustrate some of these issues.

The employment data we have assembled cover California and the US as a whole, for manufacturing and service industries (to 3-digit, and in some cases to 4-digit, SIC codes), going back annually to at least 1987. We have used three primary sources of employment data, each with its own advantages and disadvantages.

**The Annual Survey of Manufacturers (ASM)** **(**from the Bureau of the Census)

This source covers only the manufacturing sector (SIC codes 20 to 39, to the 3-digit level), but is also the only source to provide consistent data on sales for the same SIC codes. In addition, every 5 years, a complete _Census of Manufacturers_ is taken of all manufacturing firms, the most recent year being 1992. The _Census_ is a primary source of 4-digit SIC code data for employment and sales. Although the annual survey uses only a partial sample, the data are calibrated with the most recent census data in order to provide estimates for the universe of all manufacturing firms. In particular, data for California and the US from 1987 to 1994 are now available on a data CD, and new and revised data for 1994 and 1995 have just been released by the Department of Commerce.

Some ASM data are not released at the California state level, in order to protect the confidentiality of individual firms. However, we have developed a statistical methodology to fill in the gaps (see Appendix 1).

We use the ASM data to analyze employment in California's manufacturing sector. The California ASM data are shown in Tables 1A and 1.B for 2-digit and 3-digit SIC codes respectively. The cells shown in bold represent the cases where our algorithm has been used to compute missing values, as described in Appendix 1.   The line labeled "auxiliaries," near the bottom of each table, refers to manufacturing employment in auxiliary service centers (such as research labs or garages) that cannot be identified with a single SIC code.

**Country Business Patterns (CBP)** (from the Bureau of the Census)

This source provides the most complete set of employment data for both the manufacturing and service sectors, often to the level of 4-digit SIC codes, for the US and California. The data are similar to the Annual Survey of Manufacturers for the manufacturing

sector and to the Current Establishment Survey (see below) for the service sectors. The major drawback is that data are currently available only to 1994. Table 2 shows the US and California CBP employment data to the 4-digit level for computer manufacturing (SIC code 357) and computer services (SIC code 737). It is noteworthy that employment in the computer services industry exceeds employment in the computer manufacturing industry by a ratio of more than 4 to 1 at the US level and almost 3 to 1 at the California level.

***The Current Establishment Survey* (CES)** (from the Bureau of Labor Statistics and California's Economic Development Department)

For the US, this source covers both the manufacturing sector and service sector (SIC codes 70 to 89, to at least the 3-digit level). However, at the California state level, only a few series are available at the 3-digit level. Annual data are currently available through 1996.

Table 3 provides a comparison of the CES, CBP, and ASM data for the manufacturing and service industries, including computer manufacturing (SIC code 357) and computer services (SIC code 737). In most cases (total US manufacturing is an example), the values from the 3 sources are very similar. The differences that do exist are due to alternative definitions, collection methods, and revisions. A very large difference occurs, however, in the computer manufacturing sector (SIC code 357), where the CES data are substantially higher for both the US and California. This is due to different SIC code definitions, and to the fact that the CES data include auxiliary employment in the SIC code. For the computer service industry (SIC code 737), there are no data from the Annual Survey of Manufacturers by definition; and no California data are available from the CES.

Location Quotients for California Employment

The employment structure of California and the US can be compared with location quotients for each industry (SIC code). The location quotient is the ratio between the proportion of California's employment in SIC code i and the proportion of US employment in the same SIC code. The location quotient for industry i is thus defined as:

$$
(1) \quad LQ_i = \frac{E_i^{CA} / E_T^{CA}}{E_i^{US} / E_T^{US}}
$$

where:

E = employment,
CA = California,
US = United States
i = SIC code i,
T = total for all SIC codes.

The location quotients for California's manufacturing employment (based on the ASM data) are shown in Tables 4.A and 4.B for 2-digit and 3-digit SIC codes respectively. Starting with the 2-digit data in Table 4.A, most of the values are relatively stable over the years, although some trends are evident (such as the growth of Apparel (code 23) and the decline in Transportation equipment, (code 37)). The rank order based on the 1995 value is shown in column 2. Scientific instruments (code 38), Apparel (code 23), and Electronics (code 36) are the most important industries, each with an employment share more than 50% above the US share. Tobacco (code 21), Textile mills (code 22), and Primary metals (code 33) are the least important industries, each with an employment share less than 50% of the US share. In particular, there appears to be no tobacco manufacturing (code 21) activity in California.

The 3-digit location quotients, in Table 4.B, follow the same pattern as the 2-digit aggregates, although there can be substantial variation within each group of 3-digit codes. For example, the 2-digit location quotient for food products (code 20) in 1995 is just slightly above 1.0, while the 3-digit location quotient for Meat products (code 201) is only 0.40 and the location quotient for Canned food products (code 203) is 2.14. Similarly, the 2-digit location quotient for transportation equipment (code 37) is about 1.0, but the quotient for Missiles and space vehicles (code 376) is over 4.7, while the quotient for motor vehicles (code 371) is 0.37.

Table 4.C provides a separate listing of the 3-digit location quotient data based on the 1995 rank. This shows, for example, that Missiles and space vehicles (code 376) has the highest location quotient. It is instructive and perhaps surprising that Luggage (code 316) and Women's outerwear (code 233) also have very high location quotients. At the other end, several Tobacco and Fabric mill codes show little or no California manufacturing employment.

The location quotients for California's computer manufacturing employment and service employment at the 4-digit level (based on the CBP data) are shown in Table 5. The overall location quotient in 1994 for California's computer manufacturing (SIC 357) is 2.63, and California is most intensive in computers, storage devices, and terminals. For computer services (SIC code 737), the overall location quotient in 1994 for California is 1.36, and California is most intensive in prepackaged software.

3   Sales Data

Sales data are available for manufacturing firms from the *Annual Survey of Manufacturers* (ASM), the same source just described for employment. We have constructed an algorithm that allocates values for missing California data, in the same manner described for the

employment data (see Appendix 1). The associated *Census of Manufacturers,* taken every 5 years, is a source of 4-digit SIC code sales data that are otherwise not available.

The California data for sales are shown for Tables 6.A and 6.B for 2-digit and 3-digit SIC codes respectively. These data are formally referred to as the "value of shipments" in the ASM reports and are measured in million of current dollars. These shipments include reshipments of the same goods between different producers, and therefore will exceed, to an unknown amount, the industry-level data for value added that are provided in the state-level national income and product accounts. The cells shown in bold represent the cases where our algorithm has been used to compute missing values, as described in Appendix 1.

Sales data are available for service firms, but only at the US level, from the *Annual Survey of Services* (ASS), a new survey from the Department of Commerce. The data are illustrated in Table 7. This table shows that nationwide sales in 1995 for computer services (SIC code 737) were over $152 billion. This is almost 5 times as large as the sales in 1995 for computer manufacturing (SIC code 357), shown in Table 5.B.

4   International Trade Data

For the United States, detailed international trade data (imports and exports) are available for manufactured goods, while some aggregated data are available for traded services. For California, only derived data are available for goods exports and no data are available for either imported goods or for traded services. In this section, we first describe the available international trade data and then describe a methodology that could be applied to generate estimates of California's imported goods.

United States International Trade in Manufactured Goods

US international trade data for 2-digit and 3-digit SIC codes are currently being revised by the Department of Commerce, but preliminary estimates are available. Custom's Reports are the basic source of all US trade data for goods. These reports, however, do not classify US international trade data by industry with the same SIC codes that are used for the *Annual Survey of Manufacturers*. Therefore, there is the need to "harmonize" the industry coding. This is being carried out separately by the International Trade Administration and by a project at the National Bureau of Economic Research headed by Professor Rob Feenstra. We focus here on the more recent data, released by the International Trade Administration, which classifies imports and exports by 2-digit and 3-digit SIC codes from 1989 to 1995.

International Trade Shares for the US

The trade share for a SIC code is defined as the amount of international trade (exports, imports, or net exports) relative to the net domestic supply of the same good. In particular, the gross export (GXS), import (GIS), and net export (NXS) shares for industry i are written as:

$$(2a) \quad GXS_i = \frac{X_i}{S_i + I_i - X_i}$$

$$(2b) \quad GIS_i = \frac{I_i}{S_i + I_i - X_i}$$

$$(2c) \quad NXS_i = \frac{X_i - I_i}{S_i + I_i - X_i}$$

where:

GXS_i = export share for SIC code i,
GIS_i = import share for SIC code i,
NXS_i = net export share for SIC code i,
I_i = imports for code i,
X_i = exports for code i,
S_i = domestic shipments for code i.

In each equation, the denominator is the net domestic supply in industry i, consisting of domestic shipments plus imports minus exports. The export (GXS) and import (GIS) shares represent the "gross" level of trade, whereas the net export share (NXS) shows the net effect of exports minus imports. In many cases, the positive impacts of exports are exactly offset by the negative effects of imports, so that the *net* export share is the best (and simplest) measure to use when evaluating the effects of international trade on the economy.   When this is not true, however, it will be preferable to use the *gross* trade shares. Tables 8.A and 8.B show gross export shares at the 2-digit and 3-digit SIC code levels respectively. Tables 9.A and 9.B show gross import shares at the 2-digit and 3-digit SIC code levels respectively. Table 8.A shows that SIC codes 35 to 38 represent 4 of the 5 highest export shares for the US. At the same time, Table 9.A shows that SIC codes 35 to 38 also represent 4 of the 7 highest import shares for the US. This reflects the intensive international trade patterns on both the export and import sides for the "high tech" sectors, which is why the net export shares are often a more useful measure for evaluating the effect of international trade on the economy.

Net export shares for the US from 1989 to 1995 are shown in Tables 10.A and 10.B for 2-digit and 3-digit SIC codes respectively. Starting with the 2-digit codes, the negative values in the last row of Table 10.A show that, for all US manufacturing, there has been a continuing trade deficit (imports exceeding exports). In 1995, the trade deficit for manufactured goods represented about 3.4% of US domestic demand for all manufactured goods. Among the 2-digit industries, only 5 (food, tobacco, printing, chemicals, and instruments) had a positive trade balance in 1995. Not surprisingly, the US had the largest trade deficits for apparel (SIC 23) and leather (SIC 31). Perhaps more surprisingly, the trade deficit in Electronics (SIC 36) represented almost 8% of US demand for these products.

Table 10.C shows the net export shares for 3-digit SIC codes ranked from the highest (net exports) to the lowest (net imports).   The top US export industries are Ordnance, Aircraft engines, and Cigarettes.   The major US import industries, with imports exceeding 25% of US demand, are apparel, audio and video equipment, footwear, leather, and luggage.

United States International Trade in Services

The US trade statistics for services are much less detailed than for manufacturing. The top half of Table 11 shows the most detailed trade statistics that are available for services. The section with net exports shows that the US maintains a large trade surplus (positive net exports) in all the shown categories of services. This section also shows data for the net income derived on foreign direct investment in the computer manufacturing and computer services industries.

The importance for the California economy of the large trade surplus in travel services has been discussed in an earlier study of the Fisher Center for Real Estate and Urban Economics.[1]   The net export section of Table 11 shows that the trade surplus in computer services is likely to be equally important. Furthermore, the large trade surplus in foreign direct investment income, and in that part of royalties that could be attributed to computer software and service firms, are of the same order of magnitude as the direct net exports of computer services.

The bottom part of Table 11 shows the export, import, and net trade shares for business services and the computer services. These shares are based on the same principle used above to compute the trade shares of manufactured goods (see equation 2 above).   Although the values for the net export shares are all positive, they represent a relatively small share of total US domestic sales for services in the business and computer fields.

---

[1] See Ashok Bardhan, Cynthia Kroll and Dwight Jaffee, "The Growing Role of Foreign Trade in California's Economy," Fisher Center for Real Estate and Urban Economics, Working Paper No. 95-239.

California Trade Data

Custom's Reports can be the source for California export data, since the custom's forms require the "point of origin" of US exports. However, the data must be adjusted for various technical problems, including missing entries and trans-shipment. The federal government has not done this, but adjusted values are available from the Miser Project at the University of Massachusetts, although only at the level of 2-digit SIC codes. In contrast, Custom's Reports cannot readily be used to determine California imports because "point of destination" information is not collected on US imports.

Port data provide another possible source for California trade data (both imports and exports).[2] The state of California has 3 customs districts: Los Angeles-Long Beach, Oakland-San Francisco, and San Diego, which together cover all points of entry and exit to and from the state, connecting it to foreign destinations. Several points in the state of Nevada are also included in these districts. Since foreign trade data are not available by state in any great detail, the port data for a state are often used as a proxy for the state's actual trade figures. However, this approach obviously overestimates both the imports and the exports of a coastal/littoral state such as California. For example, a good part of the trade with Asia of the 18 states with no customs districts is likely to pass through California ports.

For the regression tests in Bardhan and Howe (1998)[3], which require only the changes in variables (not the levels), we use the port data as proxy for California's trade data. We intend, however, to use a modified version of California's port data in our future work, based on the following steps:

---

[2] Ashok Bardhan had the primary responsibility for the port data discussed in this and the following paragraph.

[3] See Bardhan, Ashok Deo and David K. Howe, "Transnational Social Networks, Transportation Costs, and the Geographic Distribution of California's Exports, Working Paper 98-262. Fisher Center for Real Estate and Urban Economics, University of California, Berkeley, 1998

1)   We will use the Miser data for California exports (X_m), and the port data for California exports (X_c), to create the separating ratio SR, SR = X_m/X_c .

2)   We will then multiply the imports through California customs districts by SR to provide estimates of California imports.

This assumes that the ratio between the Miser and the port export estimates applies as well on the import side. This requires that the geographic pattern of exports and imports be similar, and that the trade of other states not be growing at a rate dramatically differently from that of California. We may also adjust the import figures for changes in the US share of imports from Asia.

A Proposed Method for Directly Estimating International Goods Imports for California

Our project has developed another methodology that could construct California import (and export) data from the highly detailed input/output data that are produced by the *Census of Manufacturers.* We conclude this paper with a summary description of this methodology.

The basic principle of the proposed method is that California's share of total US imports for each commodity should reflect California's share of the activities that create the demand for that commodity. In particular, goods imported to the United States can be separated into two broad classes: goods imported to meet final demand (for consumption, capital investment, or government use) and goods imported as intermediate products in a production process. The following describes some of the details of this proposed estimation process.

A Basic Identity

We start with a basic identity for each industry i (SIC code i) at the national level:

(3) $F_i + U_i = P_i + M_i - X_i, \quad i = 1, 2...I$, where

$F_i =$ Final demand (C+I+G) for good i (from national income accounts),
$U_i =$ Input demand (use of good i as intermediate input in all industries (including i)).
$P_i =$ Production of good i,
$M_i =$ Imports of good i,
$X_i =$ Exports of good i.

The left hand side of equation [3] is the net domestic <u>uses</u> of good i and the right hand side is its net domestic <u>sources</u>. Data for the US are available for all the series in equation [3], although the final demand data available from the National Income Accounts (NIA) are more aggregated than the SIC code data available for the other variables in equation [3].

It is not feasible to apply equation [3] to California, because the trade variables M and X involve trade between California and other states, as well as international trade. We can compute California imports, however, using California data for production and exports, if we assume that the ratio of California imports to US imports for a particular good equals the ratio of California demand to US demand for that good. The steps are described as follows.

<u>Steps to Compute California Imports</u>

First, compute the amount of intermediate inputs used in California by applying an input/output (I/O) table to California's production levels.[4] Let u_{ij} be the amount of input i used in the production of \$1 of good j, based on the I/O table. The total amount of input i required for California production is then:

(4) $\quad U_i^{CA} = \sum_j u_{ij} P_j^{CA}, \; i=1,2,...I.$

Next, compute the amount of final demand in California by using the National Income Accounts (NIA) final demand data for California and the US to allocate California's share of US final demand across SIC codes. Specifically, let $N_j^{CA/US}$ be California's share of US final demand for group j based on the NIA data for California and the United States:

(5) $\quad N_j^{CA/US} = \dfrac{N_j^{CA}}{N_j^{US}}, j=1,2,...J.$

---

[4] It would be best, of course, to apply the California I/O table, except it is released much later than the US table.

California's final demand can be translated from the J NIA categories to the larger number of I SIC codes by applying California's NIA code shares to the US SIC code aggregates. Specifically, for each NIA group j (j = 1, 2, …J), determine all the associated SIC codes that are components of group j. We can then compute the final demand by SIC codes:

(6) $\quad F_i^{CA} = F_i^{US} * \dfrac{N_j^{CA}}{N_j^{US}}, \quad for\,each\,SIC\,code\,i\,associated\,with\,NIA,\,j = 1,2...J$

The last step is to compute the California imports. To do this, for each SIC code, we would apply the California share of total demand (final demand plus intermediate inputs) to US imports:

(7) $\quad M_i^{CA} = \dfrac{F_i^{CA} + U_i^{CA}}{F_i^{US} + U_i^{US}} M_i^{US}$

Extensions of the Procedure

The above procedure assumes that the amount of California imports depends only on total demand, and, in particular, neither on the nearness of California to the exporting country nor on the level of California production of the good . In fact, one might expect California to import more goods, all else the same, when the foreign producer is relatively close. Similarly, California might import less of a given good, all else the same, if California is itself a major producer of the good.

To implement these refinements, we would have to measure and calibrate the likely magnitude of the effects. This would require that we combine our data set of US demand, production, imports, and exports by SIC code with worldwide production of the same goods. This could be based on a series of statistical tests, including the following two.

First, we could run a cross-section regression that determines the ratio of US net trade to US demand for each good as a function of the US share of total world production of that good. Presumably, we would find that US net imports relative to demand are less for those goods where the US is an important international producer.   This relationship could then be applied to California's net imports as a function of its share of world or US trade in each good.

Second, we could run a cross-section regression that determines the ratio of US net trade to US demand for each good as a function of the proximity of the world's largest producer of that good. Presumably, we would find that US net imports relative to demand are greater for those goods where the world's largest producer is nearby. This relationship could then be applied to California net imports as a function of the proximity of world producers.

Data limitations and time constraints within the present study have precluded carrying out the computations of California imports described in this section. We hope to do so, however, in future work.

Appendix 1

Data from the *Annual Survey of Manufacturers* at the state level are not publicly released when that would disclose values for individual firms. For the state of California, over the period 1987 to 1995, this affected 2-digit SIC codes 22 and 31 and a number of 3-digit codes.[5]

However, the Department of Commerce does include the concealed values in its larger aggregates. For example, the aggregate manufacturing data for California include the values for the SIC codes 22 and 31, even though these values are not disclosed at the 2-digit level. Similarly, the 2-digit code data include the concealed values for the 3-digit sectors. Consequently, it is possible to create statistical algorithms to estimate the concealed values.

The values that had to be estimated to complete the data matrix for the California manufacturing sector between 1987 and 1995 are shown in Tables 1 and 6 for employment and sales respectively; the "A" tables refer to 2-digit SIC codes and the "B" tables refer to 3-digit SIC codes. All the cells shown in bold were concealed in the official statistical releases. The specific values shown in these cells have been computed by the following statistical algorithms (the same algorithm was applied to both the employment and sales data):

1) Values for SIC codes 22 and 31 between 1987 and 1991. In each year, the difference between the state total (called ACT) and the sum over the disclosed 2-digit codes (called SUM) was computed. The difference (ACT-SUM) for each year was then allocated between codes 22 and 31 based on the relative size of these two codes in 1992 (the first year for which full disclosure was available). For example, in 1987, the employment amount to be allocated was 23,000 (=ACT-SUM), 70% of which was placed in code 22 and 20% of which was placed in code 31.

---

[5] The 3-digit SIC codes are 221, 222, 223, 224, 226, 228, 229, 231, 237, 261, 262, 301, 302, 311, 313, 314, 315, 316, 317, 319, 325, 328, 333, 339, 373, 374, 375, 385, 387, 391, and 399.

2)   3-Digit SIC codes in 1992. The year 1992 has the fewest disclosure problems because

that was a complete census (rather than a survey) year. In particular, as shown in Tables

1B and 3B, there were undisclosed values in only 10 3-digit SIC codes, which

corresponded to 6 different 2-digit SIC codes.[6]   For each 2-digit code, the difference

was computed between the complete count for the 2-digit code (called ACT) and the sum

over the corresponding 3-digit codes for which data were available (called SUM). The

difference ACT-SUM was then allocated across the 3-digit codes for which values had

not been disclosed:

a)   For the 2-digit SIC codes 22 and 31, only a single 3-digit code was missing, so the
missing value was set directly equal to ACT-SUM.

b)   For each of the other 2-digit codes, data were not disclosed for 2 of the 3-digit
codes, so the amount of ACT-SUM had to be allocated across the two 3-digit
codes. For SIC codes 26, 30, and 37, the allocation was based on the nearest date
for which actual 3-digit level data were available. For example, in Table 1B for
code 262, there was actual employment of 2,000 in 1995. The amount of ACT-
SUM in that year was 2,300. Therefore, it could be deduced that the missing
value for code 261 in 1995 had to be 300. The ratio of codes 261 and 262 in 1995
was then used to allocate the amount of ACT-SUM in all other years of missing
data.

c)   For codes 231 and 237, data were not disclosed for any year in the sample.
Therefore, the amount ACT-SUM for code 23 was distributed between 231 and
237 based on the ratios of 231 and 237 at the U.S. level for each year.

---

[6] The SIC codes were 223, 231, 237, 261, 262, 301, 302, 315, 374, and 375.

<u>3-Digit Codes in Years Other than 1992</u>. The estimation of missing values for all the years other than 1992 was carried out in a manner similar to that just described for 1992. In particular, for each year, the difference was computed between the complete count for the 2-digit code (called ACT) and the sum over the corresponding 3-digit codes for which data were available (called SUM). The difference ACT-SUM was then allocated across the 3-digit codes for which values had not been disclosed. The algorithm used the same 3 subcases, a, b, and c, just described for 1992. In all cases, the allocation among the 3-digit codes was based on the 1992 values. For example, for codes 277 and 278, for each year 1987 to 1991, the amount ACT-SUM was allocated between 277 and 278 based on the relative values for these codes in 1992