

# This document is discoverable and free to researchers across the globe due to the work of AgEcon Search. 

## Help ensure our sustainability. Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

Papers downloaded from AgEcon Search may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.


## OCCASIONAL PAPER SERIES



The work reported herewithin contributes to the objectives of North Central Regional Project NC-194 a joint research project of state agricultural experiment stations and the U.S. Department of Agriculture


# THE UNITED NATIONS BILATERAL EXTERNAL TRADE DATA: CAN IT BE SAVED? 

MARINOS E. TSIGAS, THOMAS W. HERTEL AND JAMES K. BINKLEY

[^0]
## Abstract

Empirical research in international trade is constrained by the absence of a reliable, exhaustive data base on bilateral trade flows among regions. The United Nations data base on bilateral merchandise trade flows among countries is the most complete source of such data. However, it is notably unreliable. The purpose of this paper is to report on a statistical procedure for estimating systematic reporting biases, by region, for this particular data base. This serves two purposes. First, it offers a concrete quantification of one dimension of the reporting problem. Secondly, with these reporting biases in hand, it is possible to "adjust" the data. That is, one can create a set of bias-corrected trade flows.

The proposed procedure capitalizes on the fact that the UN bilateral trade data base contains two observations on every trade flow, at any point in time. Our statistical model seeks to explain the discrepancy between reported exports from country $i$ to country $j$ at time $T$, and reported imports by from $i$ at $T$. Systematic discrepancies are attributed to c.i.f./f.o.b. margins and reporting biases by the two countries in question. We estimate the model using trade flows among OECD countries over the period 1962-1987. Merchandise trade is grouped into eight categories.

Preliminary results indicate that the reporting of Japanese import flows and U.S. export flows are unbiased. Both exports from, and imports to Australia exhibit statistically significant reporting biases. In particular, exports are systematically underreported (by 11.5\%), while imports are systematically overreported (by 4.4\%): A similar pattern exists for New Zealand, while the opposite is true of the European Community. Canada, on the other hand, appears to overreport both exports and imports. Some of these results may be due to the simple approach which we have taken to modeling the transportation and insurance margins. Future research will attempt to improve this aspect of the model. We will also attempt to estimate reporting biases for non-OECD regions.

## INTRODUCTION

Empirical research in international trade is constrained, by the absence of reliable, exhaustive data on bilateral trade flows among regions. The United Nations data on bilateral merchandise trade among countries is the most complete source of such data. However, it is notably unreliable. The purpose of this paper is to report on a statistical procedure for estimating systematic reporting biases, by region, for this particular data base. This serves two purposes. First, the presence of significant, systematic biases is of interest for its own sake. Second, with estimates of reporting biases available, it is possible to "adjust" the data. That is, one can create a set of bias-corrected trade flows.

## PREVIOUS RESEARCH

There is a rather long history of attempts to explore, and sometimes correct for, the sources of inconsistency in world trade statistics. Parniczky provides a useful overview of the history of these attempts. He traces them back to Zuckermann (1920) and the League of Nations (1935-38), with more recent efforts to reconcile trade statistics being initiated by Canada and the U.S. (Bureau of the Census, 1970) and the U.N. Statistical Office (1974). Hiemstra and Mackie (1985) outline an ongoing effort by USDA to reconcile disaggregated agricultural trade data from the U.N.

Parniczky identifies the following major sources of inconsistency in reported trade:
(1) Time lags between the date of an export transaction and observation of the corresponding import.
(2) Differential administrative attention. In particular, he notes the incentive for governments to keep better records on items where quantitative controls or tariff revenues are involved.
(3) "Misclassification" of commodities, or discrepancies in the way the same commodity is mapped from domestic to Standard International Trade Classification (SITC) codes.
(4) Transportation and insurance costs.
(5) Transshipment. In particular he notes (p. 45) that: "Frequently the exporter is not aware of the final destination of the merchandise and the importer has multiple choices in identifying the country of provenance." Even where such uncertainties do not exist, the coexistence of two competing procedures for identifying trading partners introduces problems when a commodity is transshipped. The "general trade system" reports all goods entering (leaving) the national territory as imports (exports), whereas the "special trade system" records only those imports destined for home use, with the country of origin being the producer. Parniczky notes that while trade statisticians recommend the former procedure, the majority of countries use the latter, as it is of greater interest to trade policy makers.

Other authors have used statistical methods to explore specific
hypotheses. For example DeWulf explores the possibility of conscious underinvoicing of imports to circumvent quotas or tariffs, or overinvoicing to take advantage of rationed foreign exchange. The correlation of trade data discrepancies with the incentive to smuggle has also been examined (McDonald). Neither of these studies finds much explanatory power in these variables. Yet the evidence of persistent discrepancies in reported exports and imports for a given transaction is overwhelming--as will be shown below. Are these discrepancies purely random, or is there a systematic component to them? The object of our paper is to answer this question.

## STRUCTURE OF THE DATA

Figure 1 provides a "picture" of one year's data available from the U.N. for each of the $K$ traded commodities, from 1962 to the latest year, as reported by the $N$ countries which together exhaust the list of all reporters, or potential reporters in the data base. The top matrix ( $X_{i j}$ ) captures exports from country $i$ to $j$ as reported by $i$. The bottom matrix in figure 1 shows import flows among the $N$ countries ( $M_{i j}$ ) as reported by country $j$.

As noted above, there are numerous problems with the U.N. bilateral trade data. Countries may report inaccurately, or they may fail to report altogether. Unlike with some trade data bases, the U.N. statistical office does not attempt to estimate values for missing observations, or to correct obviously erroneous data. Thus, if we were to simply use the data in its raw form we would dramatically understate the role of some countries (e.g., the centrally planned economies--many of whom do not report at all) in international trade, and others may be given excessive importance.

In an attempt to improve upon the quality of this data, we formulate a model of the data generation process, which we subsequently estimate. In this manner we are able to bring the full time series of import and export data matrices, for all K commodities, to bear in our effort to specify a single benchmark trade data set.

Figure 1. The Structure of the United Nations External Trade Data for a Given Commodity in a Single Year


REPORTER $j=1, \ldots, N$

Reported Imports
$M_{i j}$

## A MODEL OF DATA GENERATION

We hypothesize that the value of the shipments data is measured and reported with systematic errors. At any point in time, the error free values (denoted with an asterisk) are:
(1) $M_{i j}^{k *}=\left(M_{i j}^{k}\right) \theta_{i j}^{k, m}$, and $X_{i j}^{k *}=\left(X_{i j}^{k}\right) \theta_{i j}^{k, x}$
where: $k$ is the index for commodities,
$i, j$ are indexes for origin and destination of shipment,
$M_{i j}^{k}$ is c.i.f. value of imports reported by $j$,
$X_{i, j}^{k}$ is f.o.b. value of exports reported by $i$,
$\theta_{i j}^{k, m}$ is the reporting error associated with the value of the import of commodity $k$, from country $i$, as reported by $j$, and
$\theta_{i j}^{k, x}$ is the reporting error associated with the value of the export of commodity $k$, to country $j$, as reported by $i$.

We assume that these errors are systematic, and may be modeled as being specific to the reporting region, although different for exports, than for imports. In other words:
(2) $\quad M_{i j}^{k *}=\alpha_{j}^{m} \cdot M_{i j}^{k}$, and $X_{i j}^{k *}=\alpha_{i}^{x} \cdot X_{i j}^{k}$ where $\alpha_{j}^{m}$ and $\alpha_{i}^{x}$ are systematic reporting biases associated with imports to, and exports from, countries $j$ and $i$. A value of $\alpha_{j}^{m}>1$ indicates that country $j$ systematically underreports imports, and similarly for i's reporting of exports when $\alpha_{i}^{x}>1$.

At the most disaggregate level, data on $X_{i j}^{k}$ and $M_{i j}^{k}$ are available in quantity terms. However, for purposes of the present paper, where we seek to analyze all merchandise trade simultaneously, such detail is not possible. Consequently data are available only in value terms. This means that unbiased imports and exports will generally differ as a result of transportation and insurance costs. ${ }^{1}$ Formally:
(3) $\mu_{i j}^{k}=\frac{M_{i j}^{k *}}{X_{i j}^{k *}}$,
where we expect $\mu_{i j}^{k}$ to be greater than 1 , with ( $\mu_{i j}^{k}-1$ ) representing the c.i.f.-f.o.b. margin.

Singularity prevents us from estimating both sets of reporting biases and flow-specific margins simultaneously. Thus we postulate a common value of $\mu$ for each commodity $k$, so that:
(4) $\mu^{k}=\frac{\alpha_{j}^{m}}{\alpha_{i}^{x}} \frac{M_{i j}^{k}}{X_{i j}^{k}}$, for all $i, j$.

Letting: $y_{i j}^{k}=\frac{M_{i j}^{k}}{x_{i j}^{k}}$, we may rewrite (4) in a manner which uses knowledge of the c.i.f.-f.o.b. margin and the reporting biases together to predict the ratio of reported imports to reported exports between $i$ and $j$ for a given commodity k:

$$
\begin{equation*}
y_{i j}^{k}=\mu^{k} \frac{\alpha_{i}^{x}}{\alpha_{j}^{m}}, \text { or } \ln \left(y_{i j}^{k}\right)=\ln \left(\mu^{k}\right)+\ln \left(\alpha_{i}^{x}\right)-\ln \left(\alpha_{j}^{m}\right) \tag{5}
\end{equation*}
$$

This is the model we wish to estimate, using time series data for an exhaustive grouping of merchandise trade, $k=1, \ldots, K$.

The regression model becomes:
(6). $\quad \ln y_{i j t}^{k}=\sigma+\sum_{\ell \neq \mathrm{bc}}^{\mathrm{K}} \gamma^{\ell} \mathrm{D}_{\ell}^{\mathrm{c}}+\sum_{\ell \neq \mathrm{bx}}^{\mathrm{N}} \beta_{\ell}^{\mathrm{x}} \mathrm{D}_{\ell}^{\mathrm{x}}+\sum_{\ell \neq \mathrm{bm}}^{N} \beta_{\ell}^{\mathrm{m}} D_{\ell}^{\mathrm{m}}+\epsilon_{i j t}^{\mathrm{k}}$,
where the error $\epsilon_{i j t}^{k}$ is independently and identically distributed, the $D^{\prime}$ s are indicator variables; $k=1, \ldots, K ; i$ and $j=1, \ldots, N$; and $t=1, \ldots, T$. The first set of indicator variables in (6), $D_{l}^{c}$, take on a value of 1 when $\ell=k$ and 0 otherwise. Thus $\gamma^{\ell}$ picks up the departure of the $k$ th c.i.f./f.o.b. ratio (in logarithms) from that of the ratio for the good which has been chosen as the "base" commodity and is thus a component of the intercept term $\left(\gamma^{\mathrm{bc}}\right)$. The second and third sets of indicator variables, $D_{l}^{\mathrm{X}}$ and $\mathrm{D}_{l}^{\mathrm{m}}$, take on a value of one and minus one when their indices equal $i$ and $j$, respectively, and zero otherwise. As a result, $\beta_{l}^{\mathrm{X}}$ and $\beta_{l}^{\mathrm{m}}$ measure the departure of exporter and importer biases from $\beta_{\mathrm{bx}}^{\mathrm{x}}$ and $\beta_{\mathrm{bm}}^{\mathrm{m}}$, the base reporters, which also appear in the intercept. Thus: $\sigma=\gamma^{\mathrm{bc}}+\beta_{\mathrm{bx}}^{\mathrm{x}}-\beta_{\mathrm{bm}}^{\mathrm{m}}$. Notice these effects cannot be separated by having a dummy variable for every commodity, importer and exporter, since, for example, the sum of the commodity dummies would always equal the sum of the exporter dummies.

If we happen to have selected unbiased base reporters, such that $\beta_{\mathrm{bx}}^{\mathrm{x}}=\beta_{\mathrm{bm}}^{\mathrm{m}}=0$, then the intercept simplifies to $\sigma=\gamma^{\mathrm{bc}}$ and so $\mu^{\mathrm{bc}}=e^{\sigma}$. The remaining margins may be derived as departures from this, i.e.: $\ln \mu^{k}=\sigma+$ $\gamma^{k}$ or $\mu^{k}=e^{\left(\sigma+\gamma^{k}\right)}$...Since the natural logarithm of the base biases is zero, the remaining biases are easily derived. For example: $\ln \alpha_{i}^{x}=\beta_{i}^{x}$ or $\alpha_{i}^{x}=e^{\beta_{i}^{x}}$. Of course, when $\beta_{\mathrm{bx}}^{\mathrm{x}}$, and $\beta_{\mathrm{bm}}^{\mathrm{m}} \neq 0$ then the intercept contains three terms, and we cannot extract these individual components.

Since the OLS estimates of the differences between the bias of two countries are invariant to the choice of base reporters, the effect of
choosing a different base for one of the reporter biases (e.g., changing bx to bx') may be captured by simply manipulating the fitted parameters: add $\beta^{x}$ bx' to the intercept, and subtract it from $\beta_{i}^{x}, i=1, \ldots, N$. Thus the new value of the intercept becomes:

$$
\sigma^{\prime}=\gamma^{\mathrm{bc}}+\left(\beta_{\mathrm{bx}}^{\mathrm{x}}+\beta_{\mathrm{bx}}^{\mathrm{x}}\right)-\beta_{\mathrm{bm}}^{\mathrm{m}}
$$

By systematically varying the base reporters, $b x$ and $b m$, and examining the value of $\mu^{k}=e^{\sigma}$, one can study the implications of a given choice of base reporter for the implied margins under the assumption of unbiased base reporters. This suggests one means of discriminating among base reporters, namely vary the combinations of $b x$ and $b m$ until the predicted margins match the evidence from other sources. However, this is a rather ad hoc procedure, and one might suspect that estimates of the c.i.f./f.o.b. margins from these other sources may not be free of the problems introduced by correlation with reporting biases. ${ }^{2}$ (This issue will be explored in detail below.)

## Capitalizing on f.o.b. Import Reporters

There is one idiosyncracy of the $U . N$. trade data base that may be turned to advantage. Because a few countries report imports on an f.o.b. basis, the model in (6) may be modified as follows:

$$
\begin{equation*}
\ln y_{i j t}^{k}=\sigma+\sum_{\ell=1}^{K} \gamma^{\ell} D_{\ell}^{c}+\sum_{\ell \neq b x}^{N} \beta_{\ell}^{\mathrm{x}} \mathrm{D}_{\ell}^{\mathrm{x}}+\sum_{\ell \neq \mathrm{bm}}^{N} \beta_{\ell}^{\mathrm{m}} \mathrm{D}_{\ell}^{\mathrm{m}}-\sum_{\ell \in \mathrm{fob}}^{\mathrm{K}} \gamma^{\ell} \mathrm{D}_{\ell}^{\mathrm{c}},+\epsilon_{i j t}^{k} \tag{7}
\end{equation*}
$$

As long as there exist some f.o.b. reported imports for each of the $K$ commodities, then need for a "base commodity" is eliminated. If a country reports imports on an f.o.b. basis, there is no margin, and any discrepancy between reported imports and exports is due to systematic bias or stochastic error. That is, the f.o.b. observations introduce zeroes into the data set, which break the pattern of singularity in the indicator variables, $D_{l}^{C}$. Thus the intercept is now comprised solely of the two biases: $\sigma=\beta_{\mathrm{bx}}^{\mathrm{x}}-\beta_{\mathrm{bm}}^{\mathrm{m}}$. Thus, this idiosyncratic pattern of reporting permits us: (a) to obtain
estimates of the c.i.f.-f.o.b. margin which are independent of the base reporters chosen, and (b) to test statistically for biased pairs of reporters. In particular, we are interested in testing: $H_{o}: \sigma=0$ against $H_{A}: \sigma \neq 0$. This in turn has important implications for the manner in which we aggregate reporters, which is the subject of the next section.

COUNTRY AND COMMODITY AGGREGATION
Due to the immense size of the U.N. trade data base, considerable aggregation is necessary before an operational data set may be obtained. The total number of observations in a given bilateral trade data set may be calculated as $N \times(N-1) \times K \times T \times 2$, where $N$ is the number of countries, $K$ is the number of commodities, $T$ is the length of the time series (in years) and there are two observations on each flow (i.e., reported imports and exports).

Given our underlying interest in trade modeling, we specified an exhaustive grouping of countries. Furthermore, it was important to break out the major f.o.b. reporters of imports. This, combined with a special interest in North American and Pacific trade, led to the grouping of countries or areas into the nineteen regions eight of which are single countries displayed in table 1. A complete description of regions appears in appendix table 1. The commodity aggregation scheme which we employed was based on ongoing trade research conducted in the Agricultural and Trade Analysis Division of the Economic Research Service of USDA, which provided the data for this project. It is summarized in the second part of table 1. Note that there are three natural resource-based commodities: food and agricultural products, forestry products, and commodities based on mining and resource extraction. Manufactured products, exclusive of those linked with the natural resourcebased commodities, are divided into five groups. The basic intermediate category includes capital intensive products (e.g., primary metals,
manufactured fertilizers and electricity) which are used in the further manufacture of other goods. Light industry includes products such as furniture, leather goods and clothing, and the high technology category consists of products such as scientific equipment and electrical machinery. The remaining two categories are intermediate manufactures (e.g., metal manufactures, office supplies, printing and publishing) and finished capital goods such as motor vehicles.

The data set was obtained from the U.N. tapes, and it reflects data availability as of March 1989. It includes observations over the period 19621987. After aggregation, there is a possible total of 146,848 reported trade values. From this data set we exclude 4,576 values which refer to interregional trade (e.g., exports or imports among the twelve countries comprising the European Community) which arise due to aggregating over countries. This leaves us with 142,272 trade values from which we may potentially compute 71,136 values for variable $y(=M / X)$.

Table 1. Aggregation Scheme: UN External Trade Data

| A. Country or Area Aggregation Eleven Regions | Eight Countries |
| :---: | :---: |
| Communist Asia | Australia (f.o.b. reporter) |
| Eastern Europe | Brazil |
| E.C. | Canada (f.o.b. reporter) |
| Latin America | Japan |
| Middle East and North Africa | Mexico |
| New Asian NICs | New Zealand |
| O1d Asian NICs | USA (f.o.b. reporter prior to 1974) |
| Other Southeast Asia | USSR (f.o.b. reporter) |
| Other Western Europe |  |
| South Asia | $\backslash$ |
| Subsaharan Africa |  |
| B. Commodity Aggregation |  |
| Aggregate Commodity | UNSITC Codes Included |
| Food, Agriculture and Fisheries | 00-02, 03, 04-23, 29, 41-43 |
| Basic Intermediate | $\begin{aligned} 266,267,35,52-53,55-59, & (\backslash 5595) \\ 62-64, & 66-68 \end{aligned}$ |
| Mining and Resource Extraction | 27-28, 32-34 |
| Light Industry | $61,65,82-85$ |
| Forestry Products | 24, 25 |
| Finished Capital Goods | 71, 73, 95, 96 |
| High Technology | 51, 54, 72, 86 |
| Intermediate Manufactures | 69, 81, 89 |

A CLOSER LOOK AT THE DATA
In this section we examine the data in more detail. There are several cases in which one of the trade partners does not report trade (exports or imports) for that particular commodity and year. For example, the USSR does not report at all into this data base. For reporting purposes, this in effect reduces the number of regions from nineteen to eighteen. Furthermore, there are 4,166 observations in which the exporter does not report trade (i.e., there is no value for X ) whereas the importer reported trade (i.e., there is a value for M). About $52 \%$ of these observations involve Communist Asia as reporter of exports. (The People's Republic of China is also a nonreporter.)

There are also 2,547 observations in which the importer does not report trade (i.e., there is no value for $M$ ) yet the exporter reported trade (there is a nonzero value for X). Again, about $56 \%$ of these observations involve Communist Asia as a reporter of imports. These three sets of observations represent extreme cases of the problem we are trying to address. For purposes of this paper, we exclude them from further consideration, since it is meaningless to quantify the reporting bias of a non-reporter.

Finally, there are 2,769 observations with both partners not reporting trade values. These carry no information and are excluded. The final data set consists of 54,166 observations and it covers trade in all eight commodities and all 306 (= $18 \times 17$ ) trade routes, excluding trade with the U.S.S.R.

Table 2 presents a summary of the final data set in two dimensions. Observations have been grouped according to the size of variable $y$ and reported trade value (which we take as the maximum of reported exports or imports). There are 44,051 observations with a ratio value between 0.25 and 4. We argue that anything outside of this range is an extreme value, and may warrant special attention. Of the remaining observations, 6,341 have y values

Table 2. Distribution of Observations by Size of Trade Flow and Variabley

| Ratio | Size of Max (X, M) |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 001 | 802 | 003 | 004 | 005 | 306 | 807 | 008 | 309 | 810 | 311 | Row T | Total |
| 001 | 10 | 28 | 23 | 24 | 20 | 11 | 1 |  | - |  |  | 117 |  |
| 802 | 23 | 29 | 17 | 4 | 8 | 6 |  |  |  | $\stackrel{ }{\cdot}$ |  | 87 |  |
| 803 . | 52 | 37 | 29 | 8 | 5 | 7 |  |  |  | $\stackrel{.}{ }$ |  | 138 |  |
| 004 | 52 | 44 | 43 | 22 | 14 | 14 | $\dot{3}$ | 3 | $\cdot$ | $\stackrel{.}{-}$ |  | 195 |  |
| 805 | 201 | 102 | 80 | 84 | 44 | 44 | 12 | 1 |  | $\cdot$ |  | 548 |  |
| 006 | 265 | 140 | 110 | 59 | 38 | 67 | 10 | 5 |  | . |  | 694 |  |
| 007 | 623 | 351 | 363 | 221 | 146 | 208 | 40 | 42 | $i$ | $\stackrel{\square}{\circ}$ |  | 1995 |  |
| 008 | 7811 | 6818 | 7585 | 5359 | 3952 | 7291 | 2200 | 2835 | 403 | 213 | 4 | 44051 |  |
| 009 | 999 | 610 | 473 | 271 | 1 R ? | ? 49 | 39 | 59 | 11 | 5 | 1 | 2879 |  |
| 010 | 422 | 248 | 202 | 136 | 75 | 82 | 21 | 13 | 5 | 5 |  | 1210 |  |
| 011 | 322 | 187 | 142 | 68 | 52 | 60 | 11 | 9 | 5 | 1 | - | 852 |  |
| 812 | 138 | 98 | 54 | 44 | 22 | 26 | 12 | 4 | 4 | 1 |  | 400 |  |
| 013 | 80 | 88 | 52 | 48 | 16 | 14 | 10 | 4 | 1 | 3 |  | 298 |  |
| -14 | 80 | 71 | 63 | 32 | 20 | 13 | 4 | 4 | 1 | 3 |  | 288 263 |  |
| 815 | 15 | 111 | 143 | 85 | 51 | 33 | 2 | 1 | - |  |  | 441 |  |
| Column |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Total | 11071 | 8743 | 9359 | 6445 | 4825 | 8125 | 2365 | 2778 | 425 | 227 | 5 | 54168 |  |

Definition of Classes for Max (X,M)

| 201: | 0 | $<=\operatorname{Max}(X, M)<$ | 1000 |
| :---: | :---: | :---: | :---: |
| a02: | 1000 | <= Max $(X, M)<$ | 5000 |
| a03: | 5000 | < $=$ Max $(X, M)<$ | 20000 |
| 004 : | 20000 | $<=\operatorname{Max}(X, M)<$ | 50000 |
| a 05 : | 50,000 | $<\operatorname{Max}(X, M)<$ | 100000 |
| 006 : | 100000 | $<=\operatorname{Max}(X, M)<$ | 500000 |
| a07: | 50.0000 | < $=\operatorname{Max}(X, M)<$ | 1000000 |
| a08 : | 1000000 | $\leqslant=\operatorname{Max}(X, M)<$ | 5000000 |
| 808: | 5000000 | $i=\operatorname{Max}(X, M)<$ | 10000000 |
| 810: | 10000000 | C=Max $(X, M)<$ | 50000000 |
| -11: | 50000000 | '<=Max (X,M)< | 100000000 |

Note: Exporits and Imports are measured in current US $\$ 1,000$.

Definition of Classes for Ratio of $M / X$

which are larger than four. Looking at the other dimension of table 2, there are 11,071 observations with trade values below the U.S. $\$ 1,000,000$ level. There are also five observations with trade values above the U.S. $\$ 50$ billion level. Casual observation of table 2 suggests a general tendency for the proportion of extreme values of $y$ to fall, as the size of the flow increases. Also, in the case of the largest flows, the extreme values all involve reported imports in excess of exports (i.e., y > 1).

Table 3 shows the mean of the variable $y(=M / X)$ (over all commodities and years) for each one of the 306 trade routes in the data set. Rows in table 3 represent regions which reported the export value (i.e., origin of a flow), and the columns represent regions which reported the import value (i.e., destination of a flow). (The number of flows observed, and their standard deviations are reported in a similar table in appendix table 5.) By far the largest mean values involve Communist Asia as reporter of exports. For example, the mean of variable $y$ for trade from Communist Asia to Australia is about 888. Again, this is due to the absence of the PRC as a reporter in all years, except 1984. On the other hand, trade routes among single country regions have consistently small means for $y$. The same is true for the European Community and Other Western Europe.

The information in table 3 suggests that the bias coefficients $\beta_{l}^{\mathrm{x}}$ and $\beta_{l}^{\mathrm{m}}$ in (7) consist of two components. One component captures the effect of not reporting trade at all. When this arises at the individual country level, we obtain a zero observation. These data points have been excluded. When it occurs within a region, the consequence is a dramatic value for $y$. The latter case arises for Communist Asia, due to non-reporting by the People's Republic of China. Since there are countries in these regions which report trade, these observations are retained in the data (i.e., neither $X$ nor $M$ is zero).

Table 3. Means of Variable y (=Mij/Xij) by Route

| Origin | Destination |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | ADS? | BRAZ | CA.N | COAS | EC12 | EEIJR | JAPA | LATA | MEX | MIDE | NNIC | NZ | OtSE | CtWE | ONIC | SAS I | SUES | IISA |
| AUST |  | 5.36 | 32.69 | 1.77 | 1.36 | 10.07 | 1.73 | 2.52 | 1.85 | 1.50 | 4.00 | 1.16 | 0.67 | $2.76{ }^{\circ}$ | 1.59 | 1.16 | 7.28 | 2.12 |
| BRAZ | $1.7 \dot{9}$ |  | 1.50 | 11.67 | 1.22 | - 5.42 | 30.55 | 0.83 | 0.98 | 2.51 | 4.17 | 2.87 | 1.43 | 1.66 | 1.30 | 2.58 | 4.01 | 1.11 |
| CAN | 1.02 | 1.29 |  | 0.62 | 1.08 | 0.62 | 1.22 | 0.70 | 0.88 | 0.70 | 1.26 | 1.26 | 0.97 | 1.31 | 1.08 | 1.00 | 4.27 | 1.06 |
| COM ASIA | 888.3 | 5.83 | 1225 |  | 1168 | 869.5 | 1324 | 140.7 | 76.50 | 1367 | 1062 | 156 | 313.6 | 1334 | 4488 | 511.4 | 2041 | 823.9 |
| EC12 | 1.08 | 1.18 | 1.09 | 1.19 |  | 0.47 | 1.27 | 0.87 | 1.17 | 0.76 | 1.73 | 1.49 | 1.14 | 1.01 | 1.38 | 1.00 | 0.78 | 1.12 |
| EEUR | 71.63 | 120.4 | 99.46 | 40.78 | 3.84 |  | 91.49 | 18.29 | 36.89 | 4.95 | 86.50 | 19.71 | 29.31 | 3.33 | 64.20 | 73.66 | 64.51 | 30.58 |
| JAPAN | 1.08 | 1.30 | 4.57 | 0.84 | 1.11 | 0.48 |  | 1.39 | 2.46 | 0.72 | 1.38 | 1.68 | 0.74 | 2.06 | 1.03 | 1.18 | 2.78 | 1.29 |
| LAT AMER. | 5.98 | 1.97 | 27.41 | 7.00 | 2.06 | 4.37 | 23.22 |  | 4.68 | 63.58 | 4.32 | 9.25 | 4.78 | 10.70 | 83.83 | 5.13 | 7.71 | 1.78 |
| MEX | 4.55 | 1.52 | 5.04 | 3.31 | 1.51 | 2.82 | 5.27 | 0.87 |  | 2.55 | 31.87 | 17.42 | 3.80 | 2.46 | 12.51 | 10.65 | 4.84 | 2.18 |
| MID FAST | 456 | 13.99 | 7:34 | 252 | 1.99 | 075 | 5380 | 3.09 | 1205 |  | 527 | 3.67 | 1133 | 1.72 | 11.31 | 8.5 | 692 | 14 ? |
| NEW NICS | 1.18 | 4.46 | 1.54 | 4.49 | 1.48 | 20.87 | 1.38 | 2.38 | 18.22 | 1.59 |  | 1.44 | 0.98 | 8.02 | 1.15 | 5.29 ' | 1.10 | 1.84 |
| NZ | 1.17 | 2.44 | 1.25 | 1.57 | 1.78 | 1.35 | 1.65 | 1.29 | 2.07 | 4.37 | 1.88 |  | 0.77 | 4.12 | 6.50 | 1.44 | 9.47 | 3.61 |
| OtSEASIA | 18:85 | 5.24 | 6.40 | 97.81 | 17.51 | 275.4 | 6.07 | 7.54 | 8.25 | 3.45 | 13.69 | 5.85 |  | 15.86 | 3.97 | 4.83 | 23.16 | 3.77 |
| OtWEur | 1.10 | . 1.22 | 2.81 | 2.29 | 1.03 | 0.54 | 1.33 | 0.98 | 2.14 | 0.90 | 2.28 | 1.39 | 0.94 |  | 2.66 | 0.98 | 1.35 | 2.82 |
| OLDNICs | 0.93 | 2.05 | 1:47 | 3.27 | 1.18 | 3.07 | 1.27 | 1.48 | 1.87 | 0.96 | 0.68 | 1.26 | 1.55 | 3.36 |  | 0.87 | 0.63 | 1.08 |
| SASIA | 5.70 | 29.14 | 14.16 | 1.45 | 4.74 | 69.53 | 4.07 | 7.57 | 6.22 | 0.86 | 4.98 | 4.17 | 17.32 | 5.95 | 2.90 |  | 2.45 | 74.84 |
| SUBSAH | 14.81 | 246.8 | 1391 | 5.40 | 8.05 | 477.6 | 2256 | 21.44 | 161.2 | 10.81 | 107.2 | 20.28 | 82.01 | 156.7 | 1270 | 142.4 |  | 24.73 |
| USA | 1.07 | 1.10 | 1.13 | 29.35 | 1.15 | 0.82 | 1.22 | 0.86 | 0.87 | 0.83 | 1.52 | 1.32 | 0.83 | 1.26 | 1.06 | 0.84 | 1.31 | - |

But clearly this type of bias can easily give rise to extreme values for $y$. The second component in the reporting biases in (7) captures the effect of not reporting accurate trade values. This happens with all regions, and it will be the focus of our statistical analysis below.

We hypothesize that, due to these two components, reporting biases for multi-country regions will be significantly different from reporting biases for single countries wi.th the latter being close to zero. Furthermore, we suspect that our simple model in (7) will not be able to sort out all of these effects. We thus choose to concentrate on a subset of regions. Specifically we select the European Community (EC), Other Western Europe (OWE), Australia, New Zealand, Japan, the USA, and Canada. We call this the OECD data set, although Turkey (a member of OECD) has not been included (Turkey has been aggregated into the Middle East and North Africa region). While the OECD data set does include two regions, namely the EC and OWE, we have been able to verify that most of the countries in these regions regularly report their trade data into the U.N. data base. ${ }^{3}$ Thus we believe the influence of the first source of bias, noted above, has been minimized.

Table 4 organizes the OECD data in a similar fashion to table 2 . There are 8,529 observations in this data set with about $97 \%$ of them (i.e., 8,279 observations) having $y$ values in the range from 0.25 to 4 . Yet there remain several observations (250) with inexplicably extreme values for $y$. About 76\% of these extreme value observations involve Australia (to a large extent) and New Zealand (to a lesser extent) as a trade partner. Table 5 shows the means for $y$ by route for this subset of (250) observations. (A complete listing of extreme $y$ values associated with Australia and New Zealand are presented in

Table 4. Distribution of Observations by Size of Trade Flow and Variable y for the OECD Data

| Ratio |  |  |  | Si | ze of | Max | X,M) |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | a01 | 002 | 803 | a 04 | 305 | 006 | 007 | $a 08$ | 309 | a10 | Row Total |
| 804 | 1 | - | - | - | - |  | - |  |  | - | 1 |
| 805 | 3 | . | . | . | . | - | - | - | - | . | 3 |
| 808 | 4 | 1 | 1 | 2 | i | - | - | . | . | - | 8 |
| 207 | 10 | 6 | 7 | 4 | 1 |  |  |  |  |  | 28 |
| a 08 | 752 | 857 | 1132 | 985 | 823 | 1683 | 686 | 1044 | 218 | 138 | 8278 |
| 809 | 50 | 23 | 11 | 7 | 11 | 14 | 1 | 1 | $\because$ |  | 118 |
| 810 | 13 | 4 | 2 | 4 | 19 | 5 | 8 |  | . |  | 1155 |
| a 11 | 9 | 3 | 1 | 1 | 1 | 1 | 3 | 1 | - | - | 20 |
| 31 ? | 4 |  |  |  |  |  |  | 1 |  |  | 5 |
| 013 | 2 | - | 1 | 1 | - | - | - | 1 | - | - | 5 |
| 814 | 1 | . | 2 | 1 |  | - | - | . | - | - | 4 |
| 815 | . . | . | . | 1 | 2 | - | - |  | - | - | 3 |
| Column Total | 849 | 894 | 1157 | 988 | 857 | 1683 | 698 | 1048 | 219 | 138 | 8529 |

Table 5. Mans of Variable y for Subset of Extreme Values
A. Means of Vary Low Ratios (less than 0.25) by Route Qrigin Destination

|  | CAN | EC12 | JAPA | NZ | OtWE | USA |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AUST | 0.15 |  | 0.16 |  | 0.20 | 0.25 |
| EC12 | 0.20 | . ${ }^{\circ}$ | . |  | . ${ }^{\text {a }}$ |  |
| JAPA | 0.12 | 0.07 | - | 0.13 | 0.07 |  |
| NZ | 0.12 | . | 0.11 |  | 0.08 | 0.15 |
| Otwe | 0.04 | - | . | - | . |  |

B. Means of Very High Ratios (greater than or equal to 4.00) by Route Origin

## AUST <br> CAN <br> EC12. <br> JAPA <br> NZ

| AUST | CAN | EC12 | JAPA | $N Z$ | OtWE | USA |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| - | 311.10 | 4.43 | 9.21 |  | 8.48 | 14.36 |
| . |  |  |  | 5.01 | 4.12 |  |
| . | 4.53 | - | - | 6.42 |  |  |
| 6.50 | 74.19 |  |  | 19.51 | 12.32 | 5.66 |
| 4.66 | 9.67 | 8.83 | 16.12 |  | 17.93 | 35.43 |
| . | 59.47 | . | 11.81 | 20.77 |  | 37.08 |

appendix tables 2 and 3.) It is not clear why this happens but distance does not appear to play a role. For example the mean of $y$ for trade from New Zealand to Other Western Europe is 4.12 , whereas the mean of $y$ for trade from Other Western Europe to New Zealand is only 1.39 (see table 3). The presence of these extreme values will show up in the statistical results below, at which point we will provide further discussion.

## STATISTICAL RESULTS

In this section, we report on the results obtained by estimating the model given in (7), using the OECD data set. Two sets of issues are explored. First, we test the null hypothesis that all OECD reporting pairs are unbiased. Since we do find statistically significant evidence of biased reporting, we proceed to compute point estimates of each region's import and export reporting biases. The second issue to be examined relates to the estimated c.i.f.-f.o.b. margins. Here we examine whether the model proposed above generates different estimates than does the customary approach of simply summing imports and exports and considering their ratio.

## Testing for Unbiased Reporting

Table 6 reports the estimates, $\hat{\sigma}$, along with the associated t statistics (second entry), for each of the 49 combinations of base reporters. An asterisk indicates that the associated intercept estimate is significantly different from zero, thus causing us to reject $H_{0}: \sigma=0$, for that pair of reporters. This occurs in 26 of the 49 possible cases. For example, with Australia and New Zealand as base reporters of exports and imports, respectively, the estimate of $\sigma$ is 0.1376 and it is significantly different from zero. This suggests that they are not "good" base reporters.

Table 6. Intercept Estimates in Model (7) by Base Pair of Reporters.

| Exporter | Importer |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | EC12. | OWELT. | Austr. | N.Z. | Japan | USA | Canada |
| EC12 | $\begin{aligned} & -0.0824^{4} \\ & -2.3429 \end{aligned}$ | $\begin{aligned} & 0.0869^{\star} \\ & 2.5351 \end{aligned}$ | $\begin{aligned} & 0.0184 \\ & 0.9569 \end{aligned}$ | $\begin{aligned} & -0.0115 \\ & -0.3354 \end{aligned}$ | $\begin{array}{r} -0.0375 \\ -1.0856 \end{array}$ | $\begin{aligned} & -0.0169 \\ & -0.6873 \end{aligned}$ | $\begin{aligned} & 0.0530^{\star} \\ & 2.77^{\prime} 453 \end{aligned}$ |
| OWEur. | $\begin{aligned} & -0.0279 \\ & -0.8155 \end{aligned}$ | $\begin{aligned} & 0.1414^{*} \\ & 4.0069 \end{aligned}$ | $\begin{aligned} & 0.0729^{\wedge} \\ & 3.8007 \end{aligned}$ | $\begin{aligned} & 0.0430 \\ & 1.2540 \end{aligned}$ | $\begin{aligned} & 0.0171 \\ & 0.4892 \end{aligned}$ | $\begin{aligned} & 0.0378 \\ & 1.5278 \end{aligned}$ | $\begin{aligned} & 0.1076 \star \\ & 5.5701 \end{aligned}$ |
| Austr. | $\begin{aligned} & 0.0867 \\ & 1.9496 \end{aligned}$ | $\begin{aligned} & 0.2359^{*} \\ & 6.8767 \end{aligned}$ | $\begin{aligned} & 0.1674^{*} \\ & 8.0035 \end{aligned}$ | $\begin{aligned} & 0.1378^{*} \\ & 4.0086 \end{aligned}$ | $\begin{aligned} & 0.1116^{\star} \\ & 3.2632 \end{aligned}$ | $\begin{aligned} & 0.1321^{*} \\ & 5.3638 \end{aligned}$ | $\begin{gathered} 0.2021^{*} \\ 10.4413 \end{gathered}$ |
| N 7 | $\begin{aligned} & 0.0605 \\ & 1.7454 \end{aligned}$ | $\begin{aligned} & n ? 997 * \\ & 6.5708 \end{aligned}$ | $\begin{aligned} & n 161 ?^{*} \\ & 8.1935 \end{aligned}$ | $\begin{aligned} & 01314^{*} \\ & 3.6891 \end{aligned}$ | $\begin{aligned} & 0.1054^{*} \\ & 3.0402 \end{aligned}$ | $\begin{aligned} & 01259^{n} \\ & 4.9988 \end{aligned}$ | $\begin{aligned} & 01950 * \\ & 9.6905 \end{aligned}$ |
| Japan | $\begin{aligned} & -0.0594 \\ & -1.7371 \end{aligned}$ | $\begin{aligned} & 0.1099^{*} \\ & 3.2058 \end{aligned}$ | $\begin{aligned} & 0.0414^{*} \\ & 2.1547 \end{aligned}$ | $\begin{aligned} & 0.0115 \\ & 0.3350 \end{aligned}$ | $\begin{aligned} & -0.0145 \\ & -0.4108 \end{aligned}$ | $\begin{aligned} & 0.0080 \\ & 0.2470 \end{aligned}$ | $\begin{aligned} & 0.0780^{*} \\ & 3.9281 \end{aligned}$ |
| USA | $\begin{array}{r} -0.0425 \\ -1.2418 \end{array}$ | $\begin{aligned} & 0.1268 * \\ & 3.6973 \end{aligned}$ | $\begin{aligned} & 0.0583^{*} \\ & 3.0380 \end{aligned}$ | $\begin{aligned} & 0.0284 \\ & 0.8276 \end{aligned}$ | $\begin{aligned} & 0.0024 \\ & 0.0719 \end{aligned}$ | $\begin{aligned} & 0.0230 \\ & 0.8848 \end{aligned}$ | $\begin{aligned} & 0.0929^{*} \\ & 4.8121 \end{aligned}$ |
| Canada | $\begin{aligned} & -0.0979^{*} \\ & -2.8638 \end{aligned}$ | $\begin{aligned} & 0.0714^{*} \\ & 2.0836 \end{aligned}$ | $\begin{aligned} & 0.0028 \\ & 0.1500 \end{aligned}$ | $\begin{array}{r} -0.0270 \\ -0.7868 \end{array}$ | $\begin{array}{r} -0.0529 \\ -1.5485 \end{array}$ | $\begin{aligned} & -0.0324 \\ & -1.3184 \end{aligned}$ | $\begin{aligned} & 0.0375 \\ & 1.7822 \end{aligned}$ |

Notes: First-line entry shows the intercept estimate, and second-line entry shows the corresponding t-statistic.
A asterisk indicates a significant from zero estimate.

It is important to note that, while rejecting $H_{o}$ implies that the pair of reporters in question is a biased pair, failing to reject $H_{o}$ does not imply the opposite. This is due to the fact that $\hat{\sigma}=\left(\beta_{\mathrm{bx}}^{\mathrm{x}}-\beta_{\mathrm{bm}}^{\mathrm{m}}\right)=$ ( $\ln \alpha_{b x}^{x}-\ln \alpha_{b m}^{m}$ ), which can also be zero when $\alpha_{b x}^{x}=\alpha_{b m}^{m} \neq 1$. That is, if a consistent overreporter of exports is paired with a consistent overreporter of imports, the resulting biases will cancel out of the ratio used to construct y. Thus it is possible to mistakenly infer from this model that a given pair of reporters is unbiased. For this reason, it is important to consider the full set of combinations, because examining all models provides further clues as to whether a given reporter is in fact unbiased.

Examination of the columns of table 6 indicates that Other Western Europe shows up as a biased reporter of imports, regardless of the exporter with which it is paired. Canada is always found to be a biased reporter, except when paired with itself. Australia also shows up as a frequently biased importer. Similarly, row-by-row perusal of table 6 highlights the presence of Australia and New Zealand as biased reporters of exports, except when paired with the E.C.

Table 7 concentrates on the subset of 13 base reporter combinations for which the t-statistic associated with $\hat{\sigma}$ is less than one in absolute value. It reports estimates of $\sigma$, along with the 2 -standard-deviation confidence interval bracketing this estimate, and the t-statistic. It also shows the implied estimates of reporting biases [parameters $\alpha_{j}^{m}$ and $\alpha_{i}^{x}$ in (2)]. They illustrate the fact that the point estimates of each region's reporting biases depend on the reporters chosen to be the unbiased base pair. Despite this fact, there are some clear patterns in these biases. For example, Australia

Table 7. Least Significant Intercept Estimates in Model (7) with Associated Bias Estimates

| Intercept |  |  | Export Biases |  |  |  |  |  |  |  | Import Biases |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Estimate | Low | Lpper | $t$ | EC12 | OWE ${ }^{\text {d }}$ | Aus : | NZ | Jap | USA | Can | EC!2 | OWEL | Aust | NZ | Jap | USA | Can |
| -0.028 | -0.096 | 0.040 | -0.82 | 0.95 |  | 1.10 | 1.09 | 0.97 | 0.99 | 0.93 |  | 0.84 | 0.90 | 0.93 | 0.96 | 0.94 | 0.87 |
| -0.027 | -0.098 | 0.042 | -0.79 | 1.02 | 1.07 | 1.18 | 1.17 | 1.04 | 1.06 |  | 1.07 | 0.81 | 0.97 |  | 1.03 | 1.01 | 0.94 |
| -0.017 | -0.066 | 0.032 | -0. 89 |  | 1.06 | 1.16 | 1.15 | 1.02 | 1.04 | 0.98 | 1.07 | 0.90 | 0.97 | 0.99 | 1.02 |  | 0.93 |
| -0.014 | -0.085 | 0.058 | -0.41 | 0.98 | 1.03 | 1.13 | 1.13 |  | 1.02 | 0.96 | 1.05 | 0.88 | 0.95 | 0.97 |  | 0.98 | 0.91 |
| -0.012 | -0.080 | 0.057 | -0.34 |  | 1.06 | 1.16 | 1.15 | 1.02 | 1.04 | 0.98 | 1.07 | 0.91 | 0.97 |  | 1.03 | 1.01 | 0.94 |
| 0.002 | -0.088 | 0.071 | 0.07 | 0.98 | 1.01 | 1.12 | 1.11 | 0.98 |  | 0.95 | 1.05 | 0.88 | 0.95 | 0.97 |  | 0.98 | 0.91 |
| 0.003 | -0.035 | 0.041 | 0.15 | 1.02 | 1.07 | 1.18 | 1.17 | 1.04 | 1.08 |  | 1.11 | 0.93 |  | 1.03 | 1.08 | 1.04 | 0.97 |
| 0.000 | -0.043 | 0.055 | 0.25 | 0.98 | 1.. 03 | 1.13 | 1.13 | . | 1.02 | 0.96 | 1.07 | 0.80 | 0.97 | 0.98 | 1.02 |  | 0.83 |
| 0.011 | -0.057 | 0.080 | 0.34 | 0.98 | 1.03 | 1.13 | 1.13 |  | 1.02 | 0.98 | 1.07 | 0.91 | 0.97 |  | 1.03 | 1.01 | 0.94 |
| 0.017 | -0 051 | ก. $0 \times 5$ | 050 | $\bigcirc$ |  | 110 | 109 | 0.97 | $\bigcirc$ | 0.93 | 105 | 0 88 | 0.95 | 0.07 |  | $\stackrel{\square}{\square} 98$ | م 91 |
| 0.018 | -0.020 | 0.057 | 0.98 |  | 1.08 | 1.16 | 1.15 | 1.02 | 1.04 | 0.98 | 1.11 | 0.93 |  | 1.03 | 1.08 | 1.04 | 0.97 |
| 0.023 | -0.029 | 0.075 | 0.88 | 0.96 | 1.01 | 1.12 | 1.11 | 0.98 |  | 0.95 | 1.07 | 0.90 | 0.97 | 0.98 | 1.02 |  | 0.93 |
| 0.028 | -0.040 | 0.097 | 0.83 | 0.98 | 1.01 | 1.12. | 1.11 | 0.98 |  | 0.95 | 1.07 | 0.91 | 0.97 |  | 1.03 | 1.01 | 0.94 |

and New Zealand always appear as underreporters of exports. On the other hand, Australia and New Zealand are found to be overreporters of imports, unless Australia is the base reporter of imports. Furthermore, on average, the USA and Japan are the least-biased reporters.

The sixth row in table 7 corresponds to the model with both the smallest and most insignificant estimate of $\sigma$. Furthermore, it pairs USA with Japan as base export and import reporters respectively. (The full set of parameters associated with this model are provided in the appendix table 4.) The next smallest t-value appears in the seventh row in table 7, with Canada and Australia as base export and import reporters, respectively. However, it is hard to accept this given the fact that Australia shows up as biased in its reporting of imports when paired with 5 of the 7 exporters. Indeed, since the estimates of $\alpha_{\text {Australia }}^{m}$ and $\alpha_{\text {Canada }}^{x}$ shown in table 7 are roughly equal in size, we may infer that this is an instance of offsetting biases.

## Estimating Margins

Table 8 shows estimates of the margin parameters i.e. the ratios of imports to exports, corrected for bias. The first row of estimates are obtained from the regression model. (Recall that these estimates are invariant to the choice of base reporter in this model.) Seven of the eight $\gamma^{k}$ parameters underlying $\mu^{k}$ are significantly different from zero. Of these, the implied share of the traded product's value expended on transportation and insurance ranges from $5.8 \%$, in the case of capital goods, to $51.2 \%$ in the case of mining and natural resource-based products. The former category involves high-value products, which are relatively easy to ship (e.g., automobiles and trucks), while the latter involves bulky, low value products. [We hypothesize that the relatively high (25.4\%) estimated margin on high technology products may arise due to insurance costs.]

Table 8. Margin Estimates for Model (7) and Two Alternative Models.

| Mode I | Agr. | $\begin{aligned} & \text { Basic } \\ & \text { Inter } \end{aligned}$ | Cap. Goods | Forst | $\begin{aligned} & \text { High } \\ & \text { Tech } \end{aligned}$ | Inter Manuf | $\begin{aligned} & \text { Light } \\ & \text { Ind } \end{aligned}$ | $\begin{aligned} & \text { Mining } \\ & \text { Resources } \end{aligned}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Model (7) | 1.127* | 1.164* | 1.058* | 1.367** | 1.254* | 1.095* | 1.017 | 1.512* |
| No Biases | 1.178 | 1.216 | 1. 105 | 1.421 | 1.311 | 1.146 | 1.064 | 1.573 |
| No Biases ignoring | $\text { , }{ }^{1} \text { info. }$ | 1.145 | 1.057 | 1.278 | 1.237 | 1.108 | 1.038 | 1.581 |

Notes for first row:
A * denotes underlying parameter estimate is significantiy different from zero at 10x.
A. " denotes that underlying parameter estimate in first row is significantly different.from underlying parameter estimate in third row.

The statistically insignificant margin associated with light industry products presents an interesting puzzle. This product category includes items such as apparel and leather goods, which may indeed exhibit a small margin. However, when non-OECD regions are included in the sample, this margin actually becomes negative! We hypothesize that the presence of extensive tariffs and quantitative controls on these products results in a systematic tendency to underinvoice imports. This would account for the insignificant, or possibly negative margin. It is certainly an issue which deserves further exploration.

An interesting question, alluded to above, is whether or not our explicit model of data generation results in improved estimates of the c.i.f.-f.o.b. margins. The standard procedure for coming up with these margins is quite simple. Sum the total value of imports for a given commodity, and similarly for exports. Then examine the ratio of these two, possibly taking an average over the sample period. Indeed, if the bias parameters are eliminated from the regression equation, we have a model which does precisely that--for transactions involving non-f.o.b. reporters of imports. If, furthermore, we dropped the last summation in (7), that is, if no accounting is taken of the fact that f.o.b. imports include no margins, then the corresponding OLS estimates would lead to the average of $M_{i j t}^{k} / X_{i j t}^{k}$ for each of the $K$ commodities over all transactions (including observations with f.o.b. valued imports).

The second and third rows of table 8 report these "naive" estimates of $\mu^{k}$, by commodity. Comparison of the first two rows of this table indicates that omission of the bias parameters results in inflated estimates of the margins. However, none of the $\gamma^{k}$ parameters underlying the second-row $\mu^{k}$ is significantly different from its first-row counterpart at the $5 \%$ level. Nor are these estimates distinct when taken as a group, at the $5 \%$ level.

Comparison of the second and third rows in table 8 indicates that, as expected, if we ignore the fact that some countries report imports f.o.b., the subsequent margin estimate falls. Ironically, in this case the two "wrongs" tend to be offsetting and all but one of the margin parameters in row three is now closer to that in row one! However, in the case of forestry products, ignoring the fact that Canada and Australia (and the U.S. prior to 1974) are f.o.b. import reporters is a costly error and the associated third-row parameter is significantly different from its first-row counterpart. Furthermore, this difference is large enough to cause the entire set of thirdrow margin parameters in model three to be different, at the $5 \%$ level, from those in the first-row.

## SUMMARY AND CONCLUSIONS

Trade modelers often require information on bilateral trade flows among countries/regions. The only source which supplies this information for all regions, at a significant level of commodity disaggregation is the United Nations external trade data base. However, researchers are reluctant to use this data set because of its notoriety for data discrepancies. Examination of the eight commodity-19 region data set used in this paper does little to reassure such researchers. Differences due to f.o.b. and c.i.f. valuations account for some of these discrepancies. Shipping lags, misclassification of commodities and countries, and smuggling have been suggested as other contributing factors.

In this paper, we describe and implement a statistical methodology which provides systematic estimates of country-specific reporting biases and commodity-specific c.i.f./f.o.b. margins. Among OECD countries, Australia and New Zealand stand out as biased reporters of trade data, particularly with regard to exports. The estimated model also suggests that the least biased
pair of reporters involves the combination of Japan for imports and the U.S. for exports.

In summary, we believe that this approach to modeling systematic trade reporting biases offers some promise for "saving" the U.N. bilateral trade data set. For example, the estimated biases may be applied to the raw data to obtain "bias-corrected" trade flows. Furthermore, they suggest a relative ranking of reporters for purposes of reconciling the remaining discrepancies in reported trade. In particular one might assign priority to the data as reported by the reporter with the lower systematic bias. Ultimately these discrepancies in reported imports and exports must be reconciled if this data base is to be used for trade modeling purposes.

Bureau of Census, United States Department of Commerce and Statistics Canada (1970). The Reconciliation of U.S.-Canada Trade Statistics.

DeWulf, L. (1981). "Statistical Analysis of Under- and Overinvoicing of Imports," Journal of Development Economics 8:303-323.

League of Nations. International Trade in Certain Raw Materials and Foodstuffs by Countries of Origin and Destination, Geneva. Four volumes: 1935, 1936, 1937, 1938.

Hiemstra, S.W. and A.B. Mackie (1985). Methods of Reconciling World Trade Statistics. FAS Report No. 217, Washington, D.C.:USDA.

McDonald, D.C. (1985). "Trade Data Discrepancies and the Incentive to Smuggle: An Empirical Analysis," IMF Staff Papers 32:668-692.

Parniczky, G. (1980). "On the Inconsistency of World Trade Statistics," International Statistical Review, 48:43-48.

United Nations (1974). International Trade Reconciliation Study. E/CN. 3/454.
United Nations. Statistical Papers: Commodity Trade Statistics. Statistical Office of the United Nations, New York, several issues.

United Nations (1986). International Trade Statistics Yearbook. Statistical Office of the United Nations, New York.

United Nations (1989). External Trade Data. Data available in magnetic tapes. Statistical Office of the United Nations, New York.

Zuckermann, S. (1920). Statistischer Atlas zum Welthandel. Berlin.

## Footnotes

1
A subset of countries reports imports on an f.o.b. basis (see table 1 and appendix table 1). This information greatly facilitates the estimation of reporting biases, as will be shown below.

2 For example, if exporters of a given commodity tend to underreport exports, then the inferred value of $\mu$ will be excessively large.
3
The countries/areas which do not report are: Austria and Finland did not report data for 1962; Greenland did not report data for 1962-75; and Andora and Gibraltar have never reported data.

Appendix Table 1 . Countries or Areas Comprising Regions

## 1. Subsaharan Africa

Angola, Benin, Botswana, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Congo, Djibouti, Equatorial Guinea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Ivory Coast, Kenya, Lesotho, Liberia, Madagasgar, Malawi, Mali, Mauritania, Mauritius, Niger, Nigeria, Republic of South Africa, Reunion, Rwanda, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, St. Helena, Sudan, Swaziland, Tanzania, Togo, Uganda, Upper Volta, Zaire, Zambia, Zanzibar-Pemba, Zimbabwe.
2. Latin America (excluding Mexico and Brazil)

Antigua and Barbuda, Argentina, Bahamas, Barbados, Belize, Bermuda, Bolivia, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, El Salvador, Falkland Islands, French Guiana, Grenada, Guadeloupe, Guatemala, Guyana, Haiti, Honduras, Jamaica, Martinique, Montserrat, Netherlands Antilles, Nicaragua, Panama, Paraguay, Peru, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Suriname, Trinidad and Tobago, Uruguay, U.S. Virgin Islands, Venezuela.
3. Middle East and North Africa

Algeria, Bahrain, Cyprus, Egypt, Gaza, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Malta, Morocco, Mozambique, Oman, Qatar, Saudi Arabia, Syria, Tunisia, Turkey, United Arab Emirates, Democratic Yemen, Yemen.
4. South Asia

Afganistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, Sikkim, Sri Lanka.
5. Old Asian NICs

Hong Kong, Singapore, South Korea, Taiwan.
6. New Asian NICs

Malaysia, Peninsula Malaysia, Sabah, Sarawak, Thailand.
7. Other Southeast Asia

American Samoa, Brunei, Christmas Island, Fiji, French Polynesia, Guam, Indonesia, Kiribati, Macau, New Caledonia, Norfolk Islands, Papua N.G., Philippines, Pitcairn Island, Ryukyu Island, Solomon Islands, Tokelau Islands, Tonga, Vanuatu, Wake Island, Wallis and Futuna, Western Samoa.
8. EC

Andorra, Belgium, Denmark, France, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, United Kingdom, West Germany.
9. Eastern Europe (f.o.b. except Hungary and Czech.)

Albania, Bulgaria, Czechoslovakia, East Germany, Hungary, Poland, -Romania, Yugoslavia.
10. Other Western Europe

Austria, Faeroe Islands, Finland, Gibraltar, Greenland, Iceland, Norway, Sweden, Switzerland.
11. Communist Asia

Burma, Kampuchea, Laos, Mongolia, North Korea, People's Republic of China, Vietnam.

Appendix Table 2. Data for Australia with Extreme Ratio Values

| ops | Aus YEAR | Orig | is e Dest. | Com | er $x$ | M | Patio |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 85 | AUSt | EC12 | Hite | 46034 | 192144 | 4.1740 |
| 2 | 86 | AUST | EC12 | HiTe | 57745 | 253419 | 4.3888 |
| 3 | 87 | AUST | EC12 | Hite | 75679 | 358298 | 4.7344 |
| 4 | 63 | AUST: | OtWE | Bsin | 267 | 1562 | 5.8502 |
| 5 | 64 | AUST | OtWE | Bsin | 270 | 1249 | 4.6259 |
| 6 | 77 | AUST | Otwe | Bsin | 1497 | 6867 | 4.5872 |
| 7 | 79 | AUST | OtWE | Bsin | 2559 | 13591 | 5.3111 |
| 8 | 80 | AUST | OtWE | Bin | 3034 | 18870 | 6.2195 |
| 9 | 82 | AUST | Otwe | Bsin | 936 | 7400 | 7.9080 |
| 10 | 67 | AUST | OtWE | Fors | 1 | 10 | 10.0000 |
| 11 | 86 | Alist | O+WF | Fors | 80 | 60 | 10.3333 |
| 12 | 68 | AUST | Otwe | Hito | 465 | 3349 | 7.2022 |
| 13 | 76 | AUST | OtWE | Hite | 4880 | 74227 | 15.2105 |
| 14 | 77 | AUST | OtWE | Hite | 4808 | 65873 | 14.2954 |
| 15 | 78 | AUST | Otwe | Hito | 4377 | 51884 | 11.8081 |
| 16 | 79 | AUST | Otwe | Hito | 8734 | 91104 | 10.4310 |
| 17 | 80 | AUST | OtWE | Hite | 19072 | 124483 | 8.5270 |
| 18 | 81 | AUSI | Otwe | hite | 11805 | 10981/ | 9.3026 |
| 19 | 82 | AUST | Otwe | Hito | 6831 | 82428 | 12.0868 |
| 20 | 83 | AUST | OtWE | Hite | 5130 | 122834 | 23.9053 |
| 21 | 84 | AUST | OtWE | Hite | 6220 | 98978 | 15.9125 |
| 22 | 85 | AUST | OtwE | Hite | 6742 | 94217 | 13.9748 |
| 23 | 86 | AUST | OtWE | Hit | 8281 | 89705 | 10.8326 |
| 24 | 87 | AUST | OtWE | Hite | 12987 | 95400 | 7.3458 |
| 25 | 82 | AUST | OtWE | Liln | 2977 | 608 | 0.2036 |
| 28 | 74 | AUST | OLWE | MnnR | 3756 | 24509 | 6.5253 |
| 27 | 75 | AUST | OtWE | MnnR | 8202 | 35742 | 5.7830 |
| 28 | 78 | AUS 1 | O+WE | MnnR | 4963 | 8407\% | 129370 |
| 29 | 77 | AUST | Otwe | Mnn | 4669 | 62743 | 13.4382 |
| 30 | 78 | AUST | OtWE | MnnR | 10954 | 84684 | 7.7309 |
| 31 | 79 | AUST | OtWE | MnnR | 10584 | 83351 | 7.8901 |
| 32 | 80 | AUst | OtWE | nn | 13135 | 104821 | 7.9803 |
| 33 | 81 | AUST | OtWE | MnnR | 14822 | 98639 | 6.8143 |
| 34 | 82 | AUST | Otwe | MnnR | 13811 | 88777 | 6.2832 |
| 35 | 83 | Ausi | UTwE | Mn | 8i84 | 84384 | - 0.025 |
| 36 | 84 | AUST | OLWE | MinnR | 16881 | 87789 | 5.2011 |
| 37 | 89 | AUST | JAPA | K | 723 | 8339 | 11.5339 |
| 38 | 80 | AUSI | JAPA | $k$ | 54892 | 7128 | 0.1299 |
| 39 | 77 | AUST | JAPA | Fors | 4 | 79 | 19.7500 |
| 40 | 81 | AUST | JAPA | Fors | 61. | 1072 | 17.5738 |
| 41 | 82 | AUST | JAPA | Fors | 370 | 1777 | 4.8027 |
| 42 | 70 | AUST | JAPA | Hito | 2046 | 23484 | 11.4882 |
| 43 | 71 | AUST | JAPA | Hite | 3888 | 36776 | 9. 4588 |
| 44 | . 72 | AUST | JAPA | Hite | 5841 | 41877 | 7.1695 |
| 45 | 76 | AUST | JAPA | Hite | 13554 | 88440 | 6.3775 |
| ${ }^{1} 6$ | 77 | AUST | JAPA | Hite | 16982 | $1335 ¢ 9$ | 7.8671 |
| 47 | 78 | AUST | JAPA | Hite | 18762 | 128309 | 6.7322 |
| 48 | 79 | AUST | JAPA | Hite | 22274 | 143388 | 6.4375 |
| 48 | 80 | AUST | JAPA | Hite | 37590 | 187500 | 4.9880 |
| 50 | 81 | AUST | JAPA | Hito | 34323. | 193252 | 5.8304 |
| 51 | 79 | AUST | JAPA | Lin | 22988 | 5447 | 0.2372 |
| 52 | 80 | aust | JAPA | Lin | 22923 | 4366 | 0.1905 |
| 53 | 81 | aust | JAPA | Lin | 19489 | 3178 | 0.1630 |
| 54 | 82 | AUST | JAPA | Lin | 22115 | 2937 | 0.1328 |
| 55 | 83 | AUst | Japa | Linn | 14798 | 2577 | 0.1741 |


|  | 84 | AUST | JAPA |  | 18709 | 2336 | 14 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 57 | 85 | AUST | JAPA | Liln | 19165 | 2603 |  |
| 58 | 62 | AUST | USA | K | 3340 | 826 | 5 |
| 59 | 70 | AUST | USA | بit | 1155 | 70771 | 17.03 |
| 60 | 71 | AUST | USA | Hite | 4922 | 70434 | 4.31 |
| 61 | 7.2 | AUST | USA | HiT | 6732 | 72873 | 10.82 |
| 62 | 76 | AUST | USA | HiT | 20559 | 300292 | 4.61 |
| 63 | 77 | AUST | USA | HiT | 24437 | 344267 | 09 |
| 64 | 78 | AUST | USA | Hite | 29188 | 452961 | 52 |
| 65 | 79 | AUST | USA | HiTe | 37158 | 529400 | 14.25 |
| 88 | 80 | AUST | USA | Hite | 42817 | 749228 | 50 |
| 67 | 81 | AUST | USA | Hit | 43487 | 885129 | 5.75 |
| 68 | 82 | AUST | USA | HiTe | 38787 | 822200 | 21.20 |
| 69 | 83 | AUST | USA | hite | 40443 | 704028 | 17.41 |
| 70 | 84 | Alist | !isa | Hite | 57914 | 79 ? 56 | 13.69 |
| 71 | 85 | AUST | USA | Hite | 52798 | 747580 | 4.16 |
| 72 | 88 | AUST | USA | HiTe | 45185 | 624807 | 3.83 |
| 73 | 87 | AUST | USA | Hite | 53252 | 613193 | 51 |
| 74 | 78 | AUST | USA | M | 27183 | 112950 |  |
| 75 | 85 | AUST | CAN | K | 6205 | 26445 | 26 |
| 78 | 75 | AUST | CAN | Fors | 5 |  | 20 |
| 77 | 88 | AUSI | CaN | Hile | 6991 | 1408 | 0.20 |
| 78 | 73 | AUST | CAN | HiTe | 31108 | 2762 |  |
| 79 | 74 | AUST | CAN | hite | 45278 | 3188 | 0.07 |
| 80 | 75 | AUST | CAN | $\mathrm{HiT}^{\text {c }}$ | 29005 | 2939 | 0.10 |
| 81 | 78 | AUST | CAN | HiTe | 3471 | 64807 | 18.67 |
| 82 | 78 | AUS 1 | CAN | HiT | 4520 | 72765 | 16.10 |
| 83 | 80 | AUST | CAN | HiTe | 7436 | 70723 | 9.51 |
| 84 | 81 | AUST | CAN | Hit | 8278 | 84789 | 10.32 |
| 85 | 82 | AUST | CAN | HiTe | 4145 | 53796 | 12.88 |
| 86 | 83 | AUST | CAN | HiTe | 5083 | 51839 | 10.20 |
| 87 | 84 | Alist | CAN | Hife | 6584 | 54176 |  |
| 88 | 85 | AUST | can | HiTe | T369 | 42205 | 5.73 |
| 89 | 88 | AUST | CAN | hite | 7853 | 80337 | 10.50 |
| 90 | 87 | AUST | can | hite | 10581 | 140532 | 13. 28 |
| 91 | 78 | AUST | CAN | Liln | 4015 | 563 | 0.14 |
| 92 | 79 | AUST | CAN | Liln | 5825 | 833 | 0.14 |
| 93 | 80 | AUST | can | Liln | 5079 | 1132 | 0.22 |
| 84 | 04 | Ausi | can |  | 257 | <<48 | 8.75 |
| 95 | 68 | AUST | can | Mnor | 2466 | 11014 | 4.45 |
| 98 | 89 | AUST | can | MnnR | 838 | 12950 | 20.36 |
| 97 | 10 | AUSI | can |  | 1343 | 44024 | 32.78 |
| 98 | 72 | AUST | CAN | MnnR | 196 | 38111 | 194.44 |
| 99 | 73 | AUST | CAN | R | 98 | 32529 | 338.84 |
| 100 | 74 | AUST | CAN |  | 46 | 48923 | 1063.54 |
| 101 | 75 | AUST | CAN | MnnR | 86 | 88380 | 1308.79 |
| 102 | 76 | AUST | CAN | MnnR | 28 | 96066 | 3430.83 |
| 103 | 77 | AUST | C |  | . 11313 | 1088 | 83 |


|  | Aust |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| OSS | YEAR | Orig | Dest | Com | ${ }^{x}$ | M | T10 |
| 1 | 88 | NZ | ausi | MnnR | 4 | 2647 | 9 |
| 2 | 84 | NZ | AUST | MnnR | 2800 | 13228 | 5.08789 |
| 3 | 86 | JAP | AUSt | Fo | 2 | 13 | 6. 50000 |

Appendix Table 3. Data for New Zealand with Extreme Ratio Values


Appendix Table 4. Regression Results for Model (7)


Parameter Estimates

| Parameter | Parameter Estimate | Standard Error | $\stackrel{t}{t}$ |
| :---: | :---: | :---: | :---: |
| Intercept | 0.002459493 | 0.03421448 | 0.072 |
| Margin Coetficients |  |  |  |
| MRG Agr | 0.11938111 | 0.03332065 | 3.583 |
| MRG Basint | 0.15222420 | 0.03334762 | 4.565 |
| MRG KGoods | 0.05598974 | 0.03334778 | 1.879 |
| MRG Forest | 0.31228117 | 0.03345890 | 8.333 |
| MRG HiTech | 0.22836446 | 0.03333139 | 6.791 |
| MRG IntMan | 0.09117839 | 0.03332605 | 2.736 |
| MRG Lilnd | 0.01709759 | 0.03332587 | 0.513 |
| M ${ }^{\text {d }}$ M MinRes | 0.41358247 | 0.03342824 | 12.372 |
| Export Bies Coefficients |  |  |  |
| XB EC12 | -0.03991743 | 0.02017980 | -1. 978 |
| XB OWEur | 0.01460673 | 0.02018572 | 0.724 |
| XB Austr | 0.10914343 | 0.02021348 | 5.400 |
| XB NZ | 0.10293993 | 0.02080329 | 4.948 |
| XE Jap | - U. U1891354 | U. UटU2U114 | -0.837 |
| XB Can | -0.05538606 | 0.02018535 | -2.744 |
| Impnrt Bias Conefticiants |  |  |  |
| MB EC12 | 0.04493848 | 0.02020452 | 2.224 |
| MB OWEUT | -0.12432388 | 0.02042134 | -6.088 |
| MB Austr | -0.05581345 | 0.03478519 | -1.604 |
| MB NZ | -0.02595301 | 0.02036578 | -1.274 |
| MB USA | -0.02053690 | 0.02398411 | -0.856 |
| MB Can | -0.09048582 | 0.03490818 | -2.592 |

Appendix Table 5. Statistics for Variable y by Trade Route
(Nanumber of observations. STDe mean standerd devistion)


## cowrinued)



Appendix Table 5. (continued)

(cowrimueo)

(cowrimeso)

This material is based in part on work supported by the U.S. Department of Agriculture, Cooperative State Research Service, under Agreement No. 89-34210-04238 and successor(s).

Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture.

Additional information on NC-194 and a complete list of project publications can be obtained from:

Executive Director, NC-194
Department of Agricultural Economics
The Ohio State University
2120 Fyffe Road
Columbus, Ohio 43210-1099
(614)292-2194


[^0]:    *The authors are research associate and associate professors, respectively, in the Department of Agricultural Economics at Purdue University, West Lafayette, Indiana 47907. They would like to thank T. Kelley White and the Agricultural Trade Analysis Division of ERS/USDA for providing the data for this project. In particular, Tom Vollrath, Sharlan Starr and Mary Wright were key collaborators in this exercise. Ed Overion provided critical information ori f.o.b. import reporters. This research was conducted under a cooperative agreement between ERS and Purduc. Research support from the NC-194 Committee, The Organization and Performance of World Food Systems is also gratefully acknowledged.

