American Economics Group, Inc.

Minnesota Sales and Use Tax Gap Project: Final Report

Prepared for

Department of Revenue, State of Minnesota

November 19, 2002

American Economics Group, Inc.
2100 M Street, Suite 810
Washington, DC 20037
202-328-1545

Minnesota Sales and Use Tax Gap Project: Final Report
November 19, 2002

The analysis and report were prepared by:

Eric Cook, Ph.D.

Charles W. de Seve, Ph.D.

Michael K. Evans, Ph.D.

(Alphabetically listed)

American Economics Group, Inc.
2100 M Street, Suite 810
Washington, DC 20037
Tel: 202-328-1545
Fax: 202-462-0594
Virginia Office 703-941-1520
Email: cwdeseve@AmericanEconomics.com
WWW.AmericanEconomics.com

TABLE OF CONTENTS

I. SUMMARY:

For the final phase of the Minnesota Sales and Use Tax Gap Project, American Economics Group, Inc. (AEG) has prepared this report for the State of Minnesota's Department of Revenue (DOR). This document summarizes the findings of the tax gap study and the methodology used to analyze data and construct databases.

The study began with an exhaustive literature search[1] and extensive interviews of DOR staff, particularly those serving in audit and research functions. Their insight and mastery of often difficult and arcane features of sales and use tax law provided AEG with significant assistance. Their input during each phase of the work and advice during the entire project have made the results robust.

AEG measured Minnesota's tax gap in detail for the year 2000 and prepared a set of microsimulation databases to enable DOR staff to continue research into additional sales and use tax issues. This report, in conjunction with separate "Status Reports" for Phases I, II and III of the project, describes how AEG constructed three levels of databases using transaction files of existing taxpayers and information from DOR audits, combined with measures of overall economic activity in Minnesota. The final result is a comprehensive, weighted sample of all taxpayers who now file and all who do not file but should. Detail for each taxpayer (actual and potential) includes industry classification, size of business, amount of tax payment, amount of tax liability, audit expectations, use tax liability and more.

The following sections describe how multiple years of transaction records were woven into a year 2000 Level I database; how we applied audit information to these data to construct the Level II database of all business taxpayers who filed returns; and how Input-Output technique was used to estimate the gap and construct the Level III database. Additionally, this study describes the results of a regression approach to estimate the growth in the tax gap and project it through the year 2007. Finally, the technical appendix provides a detailed "how to" of the process used to create and update the databases.

THE TOTAL GAP

The sales and use tax gap[2] occurs when there is a difference between estimated revenue from expected "full-compliance" tax collection and the actual revenue collected. The gap consists of current taxpayers who underreport and those businesses and households that should file and remit tax, but do not. Built on actual DOR transaction files that make up the Level I database, the Level II database includes imputed additional taxes that current filers should pay, but do not. The Level III database (broken into two segments: businesses and households) includes the non-

---

[1] The results of the literature search include publications and reports from academic and government source, an extensive bibliography was provided to DOR in a separate document early in the project.

[2] The analysis separates sales and use tax components when possible. It is an artifact of the estimation process that figures relating to the total gap will be more robust that the split between sales and use components.

filers and imputes the amounts they owe the state. Final estimates of the gap flow directly from analysis of these databases.

Referring to figure S1, the total Minnesota Sales and Use Tax gap—the "full-compliance" gap—for year 2000 is an estimated $451.1 million. Of this, current filers who underreported evaded paying an estimated $288.1 million (Level II data), while non-filers evaded paying $163.0 million in state-level sales and use tax (Level III data).

Of the total $163.1 million for non-filers, sales tax accounts for 27.4% or $44.7 million, while 72.6% or $118.3 million relates to use tax. Households avoiding the use tax are responsible for $74.7 million of the total non-filer gap, or 45.8%, while businesses account for $88.3 million, about 54.2%.

The total gap from E-commerce is $66.4 million and includes $30.8 million related to retail, of which $16.5 million is derived from households and $14.3 million from businesses. The remaining E-commerce includes $8.9 million in services, $22.8 million in manufacturing and wholesaling, and $3.9 million in all other activity. The $53.7 million use tax portion of the E-commerce tax gap is 80.5% of the total E-commerce gap. The additional gap from catalog sales amounts to $58.3 million.

PROJECTIONS

AEG projects that the 2000 Minnesota Sales and Use Tax gap of $451.1 million will grow to $693.1 million by 2007 (figure S2). The slight increase in the overall growth rate—from .75% to 8.1%—reflects several factors, including: E-commerce growth, the climb out of the recession and, absent additional compliance efforts, a small increasing propensity to avoid taxes. The E-commerce portion of the gap shows higher growth at a decreasing rate, falling from nearly 30% annually at present to 17.8% by 2007. The growth pattern follows an "S" curve rather than a simple compound growth curve, and

Figure S2

Projection of Total and E-Commerce Tax Gap 2000-2007
(millions of dollars)

Year	E-Commerce Tax Gap	Percent Change	Other Gap Components	Percent Change	Total	Percent Change
2000	66.5		384.6		$451.1	N/A
2001	86.2	29.7%	387.6	0.8%	$473.8	5.0%
2002	110.8	28.4%	392.9	1.4%	$503.7	6.3%
2003	133.9	20.8%	397.9	1.3%	$531.8	5.6%
2004	161.4	20.6%	403.4	1.4%	$564.9	6.2%
2005	192.3	19.1%	409.4	1.5%	$601.7	6.5%
2006	228.7	18.9%	416.1	1.6%	$644.8	7.2%
2007	269.4	17.8%	423.6	1.8%	$693.1	7.5%

Figure S1 - Total Tax Gap by Components: 2000

Industry	Under-Reporting Sales Tax Gap (Level II)			Non-Filer Sales Tax Gap (Level III)			Total Non-Filers Sales and Use Tax Gap (Level II & III)
	Sales Tax Gap	Use Tax Gap	Total Level II Gap	Sales Tax Gap	Use Tax Gap	Total Level III Gap	
Agriculture	887,733	2,331,739	3,219,472	141,603	262,825	404,428	3,623,900
Mining	41,038	1,218,197	1,259,235	26,159	304,055	330,214	1,589,449
Construction	3,882,293	8,495,327	12,377,620	746,709	2,345,567	3,092,276	15,469,896
Manuf. & Wh'sale Trade	32,390,724	46,605,981	78,996,705	12,697,435	17,490,782	30,188,217	109,184,922
Transportation & Utilities	4,121,016	9,613,383	13,734,399	118,048	1,001,024	1,119,072	14,853,471
Retail Trade	57,780,212	39,595,849	97,376,061	19,510,193	10,558,268	30,068,461	127,444,522
FIRE	361,488	7,683,993	8,045,481	-	-	-	8,045,481
Services	35,391,020	37,664,287	73,055,306	11,477,384	11,611,496	23,088,880	96,144,186
subtotal Business	134,855,524	153,208,757	288,064,280	44,717,531	43,574,017	88,291,548	376,355,828
Households	-	-	-	-	74,754,756	74,754,756	74,754,756
Total Tax Gap	134,855,524	153,208,757	$288,064,280	$44,717,531	$118,328,773	$163,046,304	$451,110,584
E-commerce Tax Gap (Included in total gap, above)							
Manufacturing & Wholesale	N/A	N/A	N/A	7,985,400	14,821,457	22,806,858	22,806,858
Retail	N/A	N/A	N/A	1,133,087	13,170,260	14,303,348	14,303,348
Services	N/A	N/A	N/A	2,155,030	6,769,393	8,924,423	8,924,423
All Other	N/A	N/A	N/A	1,668,308	2,298,102	3,966,410	3,966,410
subtotal Business	N/A	N/A	N/A	12,941,825	37,059,213	50,001,038	50,001,038
Households	N/A	N/A	N/A	-	16,496,323	16,496,323	16,496,323
Total E-commerce Tax Gap	N/A	N/A	N/A	$12,941,825	$53,555,536	$66,497,361	$66,497,361
E-commerce Total Retail (business and households)	N/A	N/A	N/A	$1,133,087	$29,666,583	$30,799,671	$30,799,671
Catalog Sales Tax Gap (Included in total gap, above)							
Households	N/A	N/A	N/A	-	$58,258,433	$58,258,433	$58,258,433

thus its rate varies. The E-commerce gap in 2007 is projected at $269.4 million, 42.5% of the gap compared to 16.9% in 2000.

THE E-COMMERCE GAP

This study reports the E-commerce tax gap as significantly less than high estimates others have made using questionable data and a more inclusive set of taxed items. To arrive at our estimate, we employed full measures of economic activity within Minnesota, detailed survey data from the U.S. Census,[3] we excluded items not taxed in Minnesota under the Sales and Use Tax (motor vehicles, clothing, etc.), and adjusted for the annual $770 use exclusion given to individuals. (Details on the exclusion are given in the report.)

From its value of $66.5 million in the year 2000, total sales and use tax loss from E-commerce sales is expected to increase to $269.4 million by 2007. Note that this is significantly less than estimates suggested by other researchers. The Fox study[4], for example, suggested a 2001 combined state and local revenue loss from E-commerce of $323.7 million in Minnesota. Fox then forecasts the gap rising to be $897.3 million in 2006 and $1,331.9 million in 2010. AEG believes these numbers are not supported by the level of business and consumer activity in Minnesota. It is common knowledge that many estimates of E-commerce and internet growth have been proven high, and apparently many studies have resorted to what appears to be excessive projections developed by Forrester.[5] In addition the U.S. General Accounting Office in 2000 studied the sales tax loss to E-commerce, and AEG's estimates are within the range of its figures for Minnesota. (The appendix contains a more complete review of the Fox

---

[3] Based upon U.S. Census' surveys. See http://www.census.gov/eos/www/papers/estatstext.pdf for a description of the surveys and for statistics on the findings. "E-commerce data were collected in four separate Census Bureau surveys. These surveys use different measures of economic activity such as value of shipments for manufacturing, sales for wholesale and retail trade, and revenues for service industries. Consequently, measures of total economic and E-commerce activity differ in concept and definition among these sectors, and should be added together with caution. The Census Bureau's E-commerce measures report the value of goods and services sold online whether over open networks such as the Internet, or over proprietary networks running systems such as Electronic Data Interchange (EDI)."

[4] See "State and Local Tax Revenue Losses from E-Ecommerce: Updated Estimates," Donald Bruce and William F. Fox, University of Tennessee, Knoxville, October 2001, published on the Web at http://www.statestudies.org/ecomreport.pdf.

[5] Forrester Research (http://www.forrester.com/home) has widely publicized its summary estimates of E-commerce. The underlying data and any analysis are available only to clients, so it is difficult to make full evaluations. Figures often cited to support extremely high internet sales apparently used $87.5 billion in national consumer E-sales for 2002 suggested by Forrester, who revised them down to $72.1 billion recently. Also, the 2007 figure was revised down from $276.6 to $217.8 billion. Forrester also estimated consumers spent $51.3 billion last year. They anticipated a 40% gain this year, whereas the Census figures show a 22% gain. It does not take many years of growth for differences of this magnitude to result in wildly different projections.

study and the basis for AEG's differences with it. It also includes a discussion of the GAO study.)

The actual Minnesota tax gap that AEG estimates rests on the total legal economic activity in the state. Unreported criminal activity such as illegal drug sales or internet pornography, for example, is included in the gap only to the extent that the income derived from it is used for subsequent legal purchases. The tax gap is based upon the full-compliance measure of all lawful economic commerce, and it represents the maximum tax attainable, if all was reported.

Full details of this AEG study are provided in the sections that follow and in the previously submitted status reports and in responses to DOR's questions. As part of Phase I work, AEG assembled an extensive bibliography of sales and use tax studies that, along with the status reports and interviews, should be considered an additional appendix to this report.

AEG has delivered to DOR the full Level I, II, and III databases, and AEG has met with DOR staff to advise them on how to construct future data updates. To further assist DOR, the technical appendix provides a "handbook" for DOR researchers on the details of database construction and tax gap methods. The techniques employed made exhaustive use of actual transaction and audit data, and the estimates for non-filers are based upon a comprehensive measure of economic activity in the state. In conclusion, AEG provided DOR with the databases, estimates and general information that constitute the most complete state-of-the-art view of the tax gap found in any research we are aware of in any state.

II. TRANSACTION AND AUDIT DATA: LEVEL I AND LEVEL II DATABASES

A. MEASURING THE TAX GAP USING MICROSIMULATION

After a thorough review of the literature and consideration of applicable methodologies, we determined that the most detailed and most consistent estimates of the tax gap can be made using microsimulation techniques. Alternative-based approaches may provide summary estimates, but it is difficult to imagine an alternative that shows the rich detail that can be gleaned from a microsimulation database.

Consider a database as the organizing media for diverse information drawn from multiple sources. Many techniques such as Input-Output analysis and econometric modeling can provide the threads to be woven into the rich microdatabase tapestry. Finally, the detail in the database stands ready for additional analysis unlike the summary tables and reported parameters of other methods. DOR can slice and dice the data to produce detailed cross-tabulations, summaries and sub-totals. Thus, AEG provided a working tool, not simply pages of tax gap numbers.

First, the approach began with the full-detailed transaction taxpayer database maintained by DOR. The methodology used the full population transaction data rather than a sample to build the frame on which to add audit detail.

Building a statistically reliable microsimulation database for the Minnesota Sales and Use Tax relies upon actual tax return information from several sources and involves a number of steps. Figure II-1 illustrates the process of beginning with transactions and adding more information to complete the full-compliance Level III database.

In the upper left of the chart, the starting population base is actual taxpayer records with several important taxpayer criteria including for each firm: 1) primary ID; 2) location number; 3) taxpayer ID; 4) SIC code; 5) county code; 6) consolidated/non-consolidated return indicator; 7) sales period (monthly, quarterly, or annual); 8) gross sales; 9) taxable sales; 10) use tax purchases; 11) sales tax; 12) use tax; 13) total sales and use tax and other data. These fields contain the requisite information to match taxpayers with audit information by using stratification according to size of firm, SIC code, sales period type and geographic region, for example.

In the chart, data from the transaction file flow into the Level I database. Construction is complex because data over several years are analyzed to create a fully comprehensive year 2000 data set. Next, the Level 1 database is augmented with audit assessment information from several years in order to produce the Level II database. (Note the flow from the upper right box.) For this purpose, audit data supplied by DOR are used both to link to actual tax returns where matched and to impute audit assessment and collection information to non-matched records. Stratification is used extensively to organize the records according to similar criteria and to match them with auditing records similarly stratified. (This is explained in more detail below.)

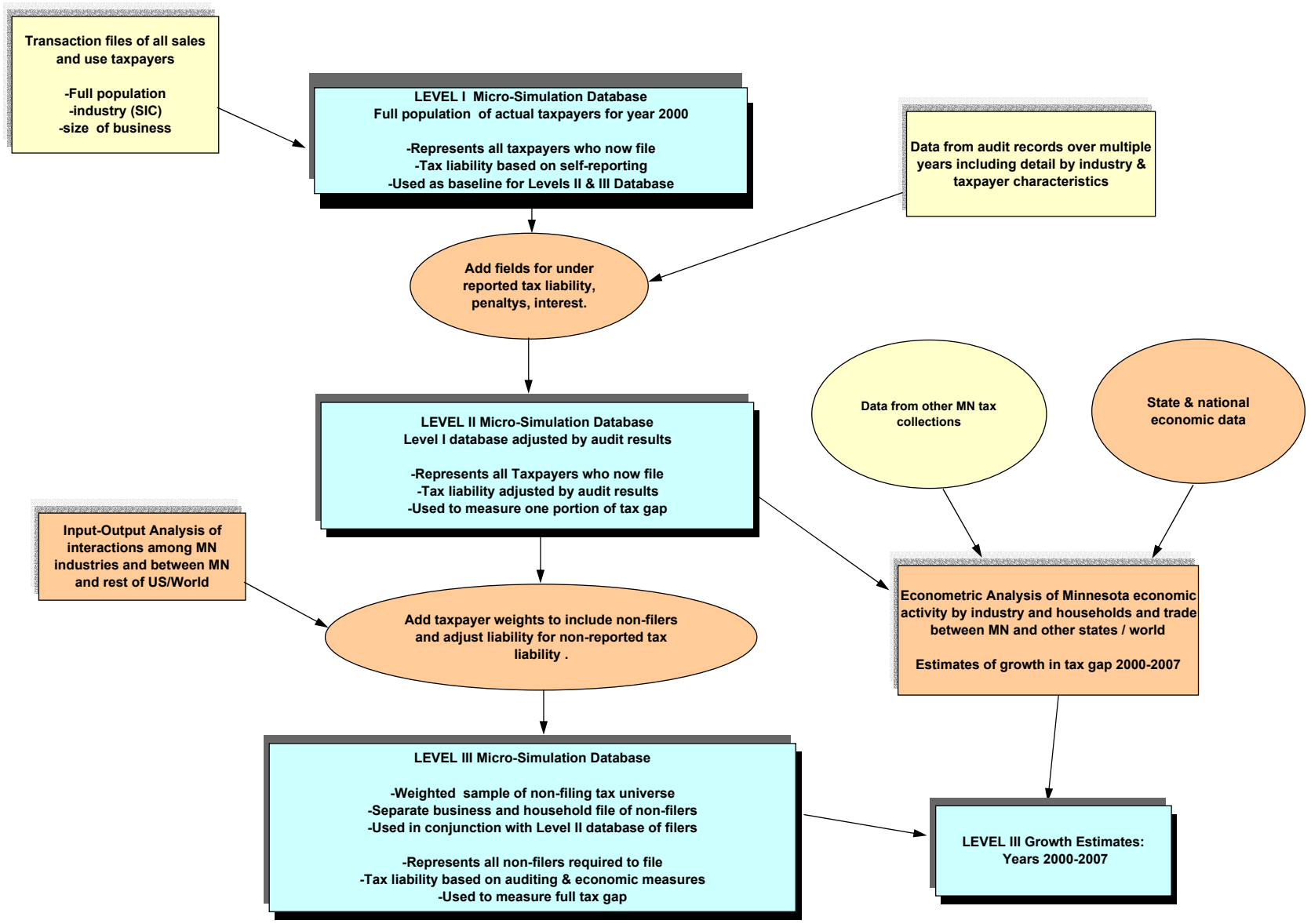
Returning to the flow chart, the Level III database requires input from the Level II database because it is presumed that every business non-filer can be described by a firm with similar characteristics that does file a return. Thus, while non-filers are not

present in the Level II database, it contains useful information to describe their characteristics and to help impute missing data.

Also flowing into the Level III database is the result of our analysis of total economic activity within Minnesota. That is shown by the Input-Output box to the left of the Level III box on the chart. By measuring the full economic interaction among industries to produce all the output of Minnesota firms and to employ all the workers from Minnesota households, we calculate the purchases required, adjust them to the state's tax code and calculate the total tax liability created in the process. (More detail on this process follows.)

On the lower right of the chart, econometric analysis is seen to produce an estimate of the dynamic movement of the tax gap over time. It is growing, and the increasing size of the gap is calculated for the years 2000-2007 using regression coupled with alternative ratio techniques, described later.

Figure II-1: Minnesota Sales and Use Tax Gap Study



B. LEVEL II DATABASE CONSTRUCTION: AUDIT EXPECTATIONS

Constructing the Level II database requires attaching audit records to transaction records. While we have a full complement of taxpayers who file returns, only portions of them are audited. By assembling audit results over many years there is a larger pool of audits to draw upon and to associate with the full transaction file to help answer the question: If all filers were audited similarly to the ones who were audited, what would be the result? By clustering (stratifying) according to taxpayer characteristics in both the transaction and audit files, audit results can be randomly drawn and associated with random records within the same cluster (stratum).

The data incorporated into Level I and Level II databases include:

-Audit Transaction Records for 1999, 2000 and 2001

-Revenue Audit Ledger Records for 1997, 1998, 1999, 2000, 2001 and 2002

-Inquiry Data for 1997, 1999 and 2000

-Transaction Data for 1995, 1996 and 1998

To begin the process of moving from Level I to Level II, a stratified random sampling of the Level I transactions database is drawn, based upon Neyman Allocation sampling techniques, which ensure a specified tolerated error level for the minimum sample size. Preliminary stratification criteria include: 1) SIC code; 2) region; 3) consolidated/non-consolidated return indicator; 4) sales period (monthly, quarterly, or annual); and 5) gross sales.

Transaction and audit records are matched using the taxpayer's primary ID as the matching key. Unmatched transaction records are associated randomly with the audit information within a given stratum, noting that a single audit record can be associated with multiple unaudited taxpayers. (The stratification and randomizing procedures are explained in the technical appendix.) The result is a Level II database containing all the Level I fields plus additional fields that show the expected assessments, penalties and interest, if each taxpayer were audited. It is important to note that some taxpayers have their taxes reduced after audit, most frequently for their failure to claim a refund for taxes paid on capital equipment purchases.

III. ESTIMATE OF TAX GAP USING INPUT-OUTPUT APPROACH

This section describes the procedure for estimating the level of economic activity in Minnesota that falls within the authority of the sales and use tax. We measure the tax gap by comparing the economic activity that creates tax liability and setting it against the actual returns and payments of tax.

A. DATA USED TO MEASURE ECONOMIC ACTIVITY AND THE SALES TAX BASE

Components of the US Bureau of Economic Analysis (BEA) Input-Output (I-O) account data for the national economy in 1992 are updated and adjusted to measure the full Minnesota economy during the year 2000. This database of economic activity in the state is constructed from several sources, including but not limited to the following:

Intermediate Business Purchases (Use) Matrix of 491-by-491 industries. Each column gives the intermediate business purchases of goods and services by a specific industry from each listed in the 491 rows. The rows, therefore, show each industry's sale of intermediate business goods and services to every other industry (and itself).

Business Capital Flow Table for New Structures and Durable Equipment Matrix contains 163 categories (columns) of new structures and durable equipment covering 64 major industry groups (rows) that purchase them.

Personal Consumption Expenditures (PCE) is a single column (vector) of consumer spending on 85 major products and services with further detailed breakdowns by SIC/NAICS codes.

These databases are "aged" or grown to the year 2000 based upon detailed national industry growth from 1992 to 2000. After these databases are set to year 2000 levels of economic activity for the nation, they are benchmarked to Minnesota levels using the various approaches discussed below. Even though the structure of industry in 1992 lies behind the data, the latest structure available, it is the more recent data that control the industry amounts.

The PCE estimates in final demand are based on data from Merchandise Line Sales (MLS). These data allocate for a particular industry all the specific goods and services it sells, which may include items predominantly sold by another industry. As a result, MLS data are a bridge between specific retail trade categories and detailed commodity categories. MLS are also used to determine the allocation of Minnesota sales to detailed commodity categories and allow comparisons to crosscheck the I-O results.

Background on BEA's I-O Intermediate "Use Table:" A commodity's (good or service) total sales or output is the sum of its sales to all industries, to be used as intermediate inputs, plus all final sales to consumers, government and business investment. An industry's total output is the sum of the intermediate inputs it buys plus the value it adds by its own productive activities (value added).

The industry's contribution to Gross Domestic Product (GDP) avoids double counting for the total output by subtracting from output all intermediate purchases, which are also included in the output of the firms that produced them. Thus, an industry's portion of GDP can be measured by the sum of its labor costs, rent, interest, dividends, etc.,

which is called its “value added” and is exclusive of its purchase of intermediary goods and services. In I-O accounts, GDP is measured either as the sum of value added by all industries or by the sum of final uses (sales) of all commodities. Across all industries these add to the same amount.

The use table gives the dollar value of goods and services purchased as intermediate products by industries. It reveals the share of commodity sales that is sold to final users. For example, some commodities, such as "apparel," are almost entirely sold to households; therefore, their demand is affected primarily by changes in the buying patterns of final users. Other commodities, such as "industrial and other chemicals," are used almost entirely as intermediate inputs; for these commodities, production is indirectly connected to final uses. As an example, "paper and allied products, except containers," is a commodity used by most industries. The largest user of this commodity is the "other printing and publishing" industry. In contrast, the commodity "metal containers," is used by only a few industries.

The estimates of intermediate inputs in the use table are based on data from various portions of the economic census. Much of these data are for broad expense categories, such as office supplies, and must be allocated to I-O commodities, such as postal services, paper and envelopes. In cases where estimates of expenses are not available, BEA uses commodity shipments and other related information to fill in the use table. (See Appendix for a listing of the Use Table industry categories, that is, the raw materials, semi-finished products and services an industry purchases from other industries.)

B. CALCULATING MINNESOTA INTERMEDIATE PURCHASES

Adjusting the US Intermediate Business Purchases matrix to current Minnesota levels of economic activity requires several steps. First, the output of each of 491 industries is allocated to Minnesota by the ratio of the state’s employment in each industry to the national total. This assumes that production functions are similar and utilization ratios are equivalent among states for a given industry. Thus, it is important to do this adjustment at the full 491-industry level because higher aggregations may have sub-industries whose weighting differs between Minnesota and the rest of the country.

County Business Patterns employment data by detailed SIC categories for the year 2000 are mapped into the 491 I-O industries for both Minnesota and the nation. Industry-specific ratios of Minnesota to US employment for year 2000 are used to benchmark the intermediate matrix columns of the aged year 2000 data matrix to Minnesota levels.

Recall that the columns of the output matrix represent industry purchases, in this case for Minnesota. The rows of the matrix represent intermediate sales of goods and services, in this case from anywhere. For the level of Minnesota output in year 2000, inputs can either come from within Minnesota or flow across state borders. However, the important observation for tax purposes is that the combined sales and use tax due Minnesota does not change if an intermediate purchase that is subject to the tax is purchased within the state or out-of-state. If it is used within the state and is taxable, then the purchase should appear in the tax base. (Dividing the base into the sales and use components is described later.)

It is a bit mind-bending to consider that the national adjustment to a Minnesota level of industrial activity requires adjustment of the matrix columns; however, the calculation of the sales tax base depends upon the purchases identified in the rows of the matrix. If the national matrix had been adjusted to Minnesota by rows instead of columns, the columns would not compute total Minnesota output. The matrix, then, would describe only that portion of Minnesota output derived from exclusively Minnesota inputs. It would not be relevant to the sales and use tax base, which at the intermediate product level depends on the place of use (Minnesota), not the place of purchase.

Thus, state business intermediate purchases are obtained when the matrix columns are benchmarked to Minnesota economic activity. The resulting rows provide estimates of sales of these intermediate goods and services.

C. CALCULATING MINNESOTA BUSINESS PURCHASES OF NEW STRUCTURES AND DURABLE EQUIPMENT

The method used to scale the US Capital Flow Table (CFT) matrix to Minnesota levels is essentially the same method used to benchmark the intermediate business purchases. The employment ratios by industry are computed for each of the industries contained in the CFT matrix. These ratios are multiplied by the US levels to obtain estimates of Minnesota purchases of new structures and durable equipment. The result of this process provides estimates of purchases in 163 categories of new structures and durable equipment by Minnesota's businesses for 64 industries (the most detailed breakdown available).

D. CALCULATING MINNESOTA PERSONAL CONSUMPTION EXPENDITURES

To scale the US personal consumption expenditures to Minnesota levels requires several steps. First, the 1992 US consumption categories are grown to year 2000 levels based upon national industry GDP growth by specific category. For example, retail trade and service categories are grown based upon the respective growth in GDP for these industries. Second, each of the 85 major US consumption estimates are scaled to Minnesota levels by multiplying each by the ratio of Minnesota population to US population for year 2000. The alternative of using employment to allocate is less connected to consumption, which is population-based.

It would be a large-scale project to account for regional differences in consumption. Minnesota's cold winters suggest a somewhat different consumer market basket than, say, in Florida. However, the differences are between Minnesota and the national average rather than just warm-weather states, for example. Furthermore, the differences only matter when the weighting would change between taxable and non-taxable products. If a Minnesota family buys more snow shovels relative to brooms than the national average, the Minnesota sales tax applies to both. Future studies could attempt to refine the consumption package, but the trade-off between the study's cost and slightly more accurate results may not be worth it.

At this point, there is sufficient information on business and household purchases within Minnesota to begin constructing a full-compliance sales and use tax base, or the revenue that would be generated to the state if everyone fully complied and paid taxes. The next step is to develop the tax rates that will be applied to the base. This is a

complex process that includes a number of other adjustments to recognize where differential tax rates apply and to eliminate excluded items and exempt purchasers from the calculated tax.

E. APPLYING MINNESOTA TAX LAW TO PURCHASES

There are different sets of tax law parameters for each of the three separate major groupings of purchases within Minnesota: 1) business intermediate purchases, with 491 categories of goods and services; 2) business purchases of producers' durable equipment, with 163 major categories; and 3) personal consumption expenditures, with 85 major categories of goods and services.

In addition, the set of tax law parameters for each of these three groupings has two components. The first component is the statutory tax rate for each good or service. The second component is the inclusion factor or "%-in-base" for a particular good or service.

Taxes and Business Intermediate Purchases: As described above, the business intermediate sales matrix contains 491 rows that represent sales of intermediate goods and services to Minnesota businesses. (The columns represent the industries making the purchases.) The first component of the tax law parameters applies an appropriate statutory tax rate to each row based upon whether or not a category is subject to the state's sales and use tax. Thus, there are 491 tax law elements that define the statutory rate for each category included in a 491-by-1 vector.

The second component of the tax law parameters is the inclusion or "%-in-base." This consists of a 491-by-491 matrix in which each element is a factor that expresses the proportion of the combination of industry and commodity (good or service) that is subject to the tax. This matrix recognizes if a particular industry is exempt from purchases of specific goods or services that are subject to tax in at least some other industries. An example of this is purchases by agricultural firms that are provided tax exemptions for certain goods otherwise subject to tax.

The values in the inclusion matrix usually take on values of zero or one. However, where a particular element represents several goods or services or where an industry contains both taxable and non-taxable entities the value would be a decimal between zero and one.

Business Purchases of New Structures and Durable Equipment: The matrix for purchases of durable equipment contains 64 industry rows that, reading across the columns, represent purchases of 163 types of structures and durable equipment by Minnesota businesses. The first component of the tax law parameters applies statutory tax rates to each of the 163 columns based upon whether the structures and durable equipment in the column are subject to the state's Sales and Use Tax law. Thus, there are 163 tax law parameters that reflect the statutory rate for each category.

As with intermediate sales, the second component of the set of tax law parameters is the inclusion or "%-in-base." This consists of a 64-by-163 matrix that recognizes the exception when a particular industry is exempt, but at least one other industry is subject to tax on that type of structure or equipment. An example of this is an equipment purchase made by a manufacturing industry that is taxable to service-industry firms. As with intermediate purchases, the element values range between zero and one. The

capital equipment tax refund is taken into account at this level and never appears as part of the tax gap, though in reality it is paid and then refunded.

Personal Consumption Expenditures: The rows in the PCE matrix reflect 85 separate categories of goods and services. The first component of the tax law parameters applies statutory tax rates to each of the 85 categories based on whether or not the item is subject to the sales and use tax. Thus, there are 85 tax law parameters that reflect the statutory rate for each category.

The second component of the tax law parameters, again the inclusion or "%-in-base," also consists of 85 elements. Their values are zero if never taxed, and one if always taxed. Decimal values between zero and one represent goods and services made to final (rather than intermediate) purchasers, which may be taxed or untaxed depending upon the purchaser's tax category. For these, an estimate is made of the % of that category subject to tax. An example is the purchase of a toaster by a convent.

F. SALES AND USE TAX GAP ESTIMATES USING THE INPUT-OUTPUT APPROACH

The Minnesota full-compliance sales and use tax base calculated using the Input-Output approach is shown in Figure III-1, adjusted for all exclusions. About 60% of the unpaid tax is owed by resident consumers and institutions making final purchases. About 40% should be paid by businesses, including both taxable intermediate purchases and structures and durable equipment. The overall Minnesota sales and use tax should have generated $4,184,285,236 for the year 2000.

Figure III-1

Full-Compliance Sales and Use Tax Estimates by Industry: Input-Output Estimates for 2000

Industry	Final Sales (Consumption)	Intermediate Business Purchases	Purchases of Structures and Equipment	Total
Agriculture		27,459,769	1,841,527	29,301,296
Mining		23,236,169	478,724	23,714,893
Construction		62,655,395	9,204,322	71,859,717
Manuf. & Wh'sale Trade		750,802,488	52,595,797	803,398,285
Transportation & Utilities	227,859,684	184,604,958	14,795,984	427,260,626
Retail Trade	1,854,280,314	127,533,043	16,210,099	1,998,023,456
FIRE		15,989,559	49,725,245	65,714,804
Services	421,288,143	282,721,428	58,604,224	762,613,795
NEC		2,398,363		2,398,363
Total	**$2,503,428,142**	**$1,477,401,172**	**$203,455,922**	**$4,184,285,236**
Percent of Total	59.8%	35.3%	4.9%	100.0%

Input-Output Estimates Compared to Actual Collections: Figure III-2 provides a comparison of the Input-Output results to actual sales and use taxes by industry.

These estimates place the overall tax gap at 10.4% of total sales and use tax collections in year 2000 at about $392.5 million. Almost half of the total tax gap, $143.9 million, derives from retail trade. However, retail's tax gap as a portion of retail's tax paid is only 7.8%, lower than the gap for most other industries.

Current filers are responsible for $288.1 million of the tax gap, underreporting an amount equal to 7.6% of what they pay (Level II database estimates). The gap caused by non-filers (from the Level III database) is $88.3 million or 2.3% of all sales and use taxes paid. Almost half of the Level III tax gap from non-filers is attributable to retail trade.

Figure III-2

Sales and Use Tax 2000: Input-Output Estimates Compared to Actuals*

Industry	Input-Output Estimate	Actual Collections	Total Tax Gap (Difference)	% Difference from Actual	Under-reporting of Filers (Level II Tax Gap)	% Difference from Actual	Non-Filers (Level III Tax Gap)	% Difference from Actual
Agriculture	29,320,297	25,696,397	3,623,900	14.1%	3,219,472	12.5%	404,428	1.6%
Mining	23,734,563	22,145,114	1,589,449	7.2%	1,259,235	5.7%	330,214	1.5%
Construction	71,905,899	56,436,003	15,469,896	27.4%	12,377,620	21.9%	3,092,276	5.5%
Manuf. & Wh'sale Trade	803,877,219	694,692,297	109,184,922	15.7%	78,996,705	11.4%	30,188,217	4.3%
Transportation & Utilities	427,739,951	412,886,480	14,853,471	3.6%	13,734,399	3.3%	1,119,072	0.3%
Retail Trade	1,982,673,817	1,855,229,295	127,444,522	6.9%	97,376,061	5.2%	30,068,461	1.6%
FIRE	65,752,494	57,707,013	8,045,481	13.9%	8,045,481	13.9%	N/A	N/A
Services	763,140,892	666,996,705	96,144,187	14.4%	73,055,306	11.0%	23,088,880	3.5%
Total	**$4,168,145,132**	**$3,791,789,304**	**$376,355,828**	**9.9%**	**$288,064,280**	**7.6%**	**$88,291,547**	**2.3%**

***Business only; does not include $74,754,756 in tax gap related to households.**

Note: The "NEC" (Not Elsewhere Classified) category for actual collections was allocated among major industry groups based upon their pro-rata share of the total sales and use tax. This was necessary to permit comparison of estimates given that the Input-Output category for "NEC" was negligible, whereas it was quite large in the actual file.

Minnesota E-Commerce Purchases: The Level III database (non-filers) provides estimates of the portions of the tax gap created by consumers and businesses. Calculations rely upon the recent Bureau of the Census survey of E-commerce and retail catalog sales. We scaled the National results from the survey to Minnesota levels based upon the ratio of the state's retail trade relative to the nation for 17 categories of goods. Estimates of the split between consumers and business were obtained using the shares in comparable categories of the Input-Output analysis. The results are shown in figure III-3.

Non-filers' use of E-commerce cost Minnesota approximately $30.8 million in sales and use tax during the year 2000. This reflects sales of a broad base, including books, electronic equipment, sporting goods, beauty aids and a range of other purchases by households and businesses. The former are responsible for $16.5 million of the gap (53.6%) and the latter $14.3 million (46.4%).

The total tax lost in E-commerce during 2000 was $66 million, including the $30.8 million noted above, $9 million related to services, $23 million to manufacturing and wholesale, and $4 million to all others. The appendix contains a commentary on the extraordinarily large and inaccurate estimates others have made.

Figure III-3

E-commerce Full Compliance Sales and Use Tax Estimates for 2000

	Households	Businesses	Total
Electronic Shopping and Mail-Order Houses			
Books and magazines	712,701	1,908,134	2,620,835
Computer hardware	1,891,393	7,104,008	8,995,401
Computer software	347,030	1,303,434	1,650,464
Drugs, health aids, and beauty aids	558,077	137,193	695,270
Electronics and appliances	1,419,844	165,497	1,585,341
Food, beer, and wine	290,112	20,976	311,088
Furniture and home furnishings	828,059	428,666	1,256,725
Music and videos	1,897,674	-	1,897,674
Office equipment and supplies	246,897	1,872,802	2,119,698
Toys, hobby goods, and games	1,117,955	58,839	1,176,794
Other merchandise	3,145,372	349,484	3,494,856
Other nonmerchandise	1,470,771	163,418	1,634,189
Traditional Retail Trade			
Motor vehicles and parts dealers	431,199	253,856	685,056
Electronics and appliance stores	726,493	84,680	811,174
Building materials and garden equipment and supplies stores	332,315	332,313	664,628
Sporting goods, hobby, book and music stores	558,200	62,022	620,222
Miscellaneous store retailers	522,230	58,025	580,256
Total	**$16,496,323**	**$14,303,348**	**$30,799,671**

IV. LEVEL III DATABASE CONSTRUCTION

A. LEVEL III DATABASE APPROACH

Recall that the Level I database consists of the total universe of all entities that file returns under the Minnesota Sales and Use Tax law. These data are derived from DOR transaction records and constructed to represent year 2000 taxpayers. Each record in the database represents one single taxpayer. The Level II database begins with Level I and adds to existing Level I record adjustments that reflect the results of DOR audits. Thus, some taxpayer records will show more tax and some less tax in Level II than the same taxpayer records in Level I. Level II has the same number of records as Level I.

The Level III database is actually the Level II database plus two add-ons contained in separate files: one for households and one for businesses. While these details could be appended to Level II data, each record stands for multiple taxpayers and is a weighted sample file rather than an all-inclusive unweighted file. If these were put together, the Level II records would require a weighting field for consistency, but it would be set to equal one for every Level II record.

Thus, we are providing the three separate components of what is the final Level III database: 1) Level II database on filers; 2) added data file on household non-filers; and 3) added data file on business non-filers.

B. LEVEL III HOUSEHOLD DATABASE

The development of the Level III household database uses the separate targets of $16.5 million for E-commerce and $58.3 million in catalog sales, discussed earlier (see figure S1). These amounts are the use taxes attributable to Minnesota households that did not file tax returns for their purchases. To create household records, we first converted the use tax estimate to total sales. Then we estimated the average sale per transaction for each of the 17 categories, relying on Consumer Expenditure Survey data. These two items are derived for each of the 17 goods and services categories shown in figure IV-1, next page. Household engaging in E-commerce and/or catalog sales had an average of 12.8 transactions per year potentially subject to the state's use tax, and we randomly distributed that number among the transaction categories.

Household data are imputed randomly using joint distributions of E-commerce and catalog purchase activity along with log-normal distributions of purchase amounts and the probability of multiple purchases in a given transaction. We constructed household tax records from the distributions, assigning weights to each record in order to limit the number of records in the database. The resulting "sample" size for the Level III household database is 60,462, with each representing six, seven or eight households as specified by the weighting variable. Note that in the final database delivered to DOR, there are separate fields that allow calculation of the $770 exclusion and separate calculations of E-commerce and catalog sales.

Figure IV-1 provides the summary detail of the Level III household database. The $16.5 million use tax for E-commerce is the same number appearing previously in figure S1,

as is the use tax for catalog sales of $58.3 million. Thus, in 2000 E-commerce and catalog sales combined should have yielded an additional $74.8 million in use tax.

Figure IV-1

Summary of E-Commerce & Catalog Sales Year 2000	
E-commerce	
Gross sales (taxable items & purchasers)	$272,207,743
Tax on Gross	$17,692,752
Excluded sales less than $770.01	$18,418,162
Net sales	$253,789,581
Tax due on Net	**$16,496,323**
Numb Hhlds	124,098
Numb transactions	1,236,654
Catalog	
Gross sales (taxable items & purchasers)	$961,347,944
Tax on Gross	$62,484,971
Excluded sales less than $770.01	$65,064,365
Net sales	$896,283,579
Tax due on Net	**$58,258,433**
Numb Hhlds	438,492
Numb transactions	4,369,926
E-commerce + Catalog	
Gross sales (taxable items & purchasers)	$1,233,555,687
Tax on Gross	$80,177,723
Excluded sales less than $770.01	$83,482,526
Net sales	$1,150,073,160
Tax due on Net	**$74,754,755**
Numb Hhlds*	438,492
Numb transactions	5,606,580
Avg. Numb trans per hld with transactions	12.8
Avg. Items per transaction	1.61
Avg transaction value	$220.02
Median transaction value	$71.04
Average item value	$136.66
Average annual spending per hhld with transactions	$2,813.18

*Virtually all E-commerce purchasers also purchase via catalogs, but not vice versa.

The 60,462 weighted household records in the database reflect combined E-commerce and catalog purchases. The gross purchases of taxable items by households subject to tax totaled $1,233.6 million. After accounting for the $770 exclusion, this became $1,150.1 million, a $83.5 million reduction. Tax due on the purchases amounted to $74.8 million of which $16.5 million was from E-commerce, as noted above.

Over the year, the households engaged in E-commerce and/or catalog purchases had 5.6 million transactions that involved an average of 1.61 items per transaction. The average transaction value was $220.02, while the median was $71.04, showing a highly skewed distribution. The value of the average item purchased was $136.66.

In general, households that engaged in E-commerce also made catalog purchases, although the reverse is not true. The 124,098 households buying over the internet accounted for 1,236,654 transactions with a gross value for the year of $272.2 million.

The 438,492 households the made catalog purchases had 4,369,926 transactions amounting to $961.3 million for the year 2000.

C. LEVEL III BUSINESS DATABASE

The Level III business database targets $87.9 million as the total tax attributable to Minnesota businesses that are non-filers evading the tax. (This is the $104.4 million total Level III in figure III-2 less the $16.5 for households in figure IV-1.) Each industry has a separate tax target, as seen in the last column in figure IV-2. Note that lumpiness and rounding in the database introduces an error of about $0.4 million in the total estimate of the gap. Thus, the table shows $88.3 million as the gap, not the $87.9 million noted above.

Assume that business non-filers have profiles similar to firms that file and that representative non-filers are found among the smallest business entities that do file. Thus, we picked random selections from the lowest decile of each major industry group randomly to create records for non-filers. We then randomly attributed to them transactions needed to sum to target totals for each industry. To limit the size of the database, each record is assigned a weight of three, thus representing three non-filing business tax entities. The split between the sales tax and the use tax is derived based upon the pattern of the filers selected. FigureIV-2 provides the results for the Level III sample of business tax returns.

Figure IV-2

Business Tax Gap of Non-Filers: Input-Output Estimates for 2000

Industry	Returns in Sample	Returns Weighted to Entire Population	Gross Sales	Sales Tax Gap	Use Tax Gap	Total Non-Filers Sales and Use Tax Gap
Agriculture	316	948	120,628,554	141,603	262,825	404,428
Mining	29	87	81,965,346	26,159	304,055	330,214
Construction	270	810	503,350,533	746,709	2,345,567	3,092,276
Manuf. & Wh'sale Trade	2,466	7,398	18,087,952,254	12,697,435	17,490,782	30,188,217
Transportation & Utilities	68	204	580,096,842	118,048	1,001,024	1,119,072
Retail Trade	6,040	18,120	4,837,599,042	19,510,193	10,558,268	30,068,461
FIRE	0	0	0	0	0	-
Services	4,665	13,995	3,164,788,747	11,477,384	11,611,496	23,088,880
Total	**13,854**	**41,562**	**$27,376,381,318**	**$44,717,531**	**$43,574,017**	**$88,291,548**

V. CHANGE IN TAX GAP OVER TIME USING REGRESSION ESTIMATES

A. INTRODUCTION

This section covers regression and alternative methods to estimate how the tax gap in Minnesota is changing over time. The regression approach is hampered to some degree by lack of detail and a sufficient history of E-commerce and other key variables. Thus, the study employs alternative measures as a way of corroborating the results. The section also includes projections of the increase in the gap through the year 2007.

AEG uses three different methods to estimate the increase in tax avoidance of Minnesota (MN) sales and use tax revenues from 1995 through 2000. These are (a) direct ratios, (b) analysis of E-sales and (c) regression estimates. We roll the approaches into a "preferred" figure that, in our judgment, represents the most likely estimate of the changing gap.

The preferred estimate of the increase in the total tax gap between 1995 and 2000 is $89 million, or about 2.4% of total sales and use taxes collected in 2000. About 40% of the increase in the tax gap occurs in retail sales: 23% in business and personal services, 18% in wholesale trade, 13% in manufacturing and 6% in construction. No measurable increase in tax avoidance is found in agriculture, mining, transportation, communications or utilities. Much of the increase represents the proliferation of E-commerce during that period.

We examined five possible sources of tax evasion. However, results indicate that there is no measurable increase in evasion during 1995-2000 from reasons (2), (3) and (4), below. Apparently the most noticeable increase occurred in E-commerce, although there are also some general increases in tax avoidance that are not linked directly to specific factors.

Sources of possible tax evasion examined:

1) Failure to pay sales and use tax on goods and services purchased over the Internet that would have otherwise been taxed

2) Increase in catalog and other non-store sales that did not take place on the Internet

3) Increase in cross-border sales with Wisconsin and possibly other neighboring states

4) Increase in Canadian sales largely because of the decline in the Canadian dollar

5) General increase in tax avoidance not due to specific causes 1 through 4

Data on Minnesota sales and use tax collections from 1995 through 2000 by detailed industry classification were aggregated into categories at the 3-digit level (SIC) for retail and wholesale trade, and 2-digit for manufacturing and services. We estimated approximately 20 regression equations for each of these four sectors, although the

coefficients representing tax avoidance were not significant in all cases, and in some cases individual industries were combined. We also estimated equations for agriculture, mining, construction, transportation, communications and utilities. Except for construction, no significant parameters for tax avoidance were determined in these subgroups.

The increase in tax avoidance since 1995 is estimated by the three separate methods noted above: direct ratios, analysis of E-sales and regression estimates. First, simple ratios are calculated to compare the growth in Minnesota tax revenues relative by total US sales to the growth in Minnesota income (or employment) relative to US employment. If tax rates did not change, revenues should be proportional to sales. Hence, a decline in the ratio of (MN tax/US sales) at the same time that the ratio of MN/US income or employment increased would suggest an increase in tax evasion.

In the second method we utilize the figures for E-sales in the retail and service sectors, and calculate the proportion of sales that take place in Minnesota. When multiplied by the appropriate tax rate, that figure indicates the loss of revenues from E-sales. Pooled cross-section time-series regressions (not shown here) reveal that approximately 80% of the total increase in tax avoidance for the 1995-2000 period is due to E-sales. Therefore this figure has been used for both retail sales and services. This method does not work for wholesale trade or manufacturing because only a small proportion of those sales represent a loss of tax revenues. Thus, separate estimates using this method are prepared only for retail sales and services, where tax avoidance is estimated to have been the largest.

The third method is to calculate the parameters of E-sales and other tax evasion terms by using regression analysis. In general, these estimates are substantially higher than the other two methods, suggesting that the pure regression equation method is incorporating trends that represent changes in tastes and preferences, and other exogenous factors that affect revenues and sales, as well as tax evasion.

B. METHOD #1: CALCULATING SIMPLE RATIOS

We first calculate the growth in Minnesota sales and use tax revenue from 1995 to 2000. This ratio is then divided by the growth in sales, shipments, or output originating in the US for the same years, depending upon the industry and data available. Sales are used for wholesale and retail trade, shipments for manufacturing, “construction put in place” for construction, farm income for agriculture, and “output originating” for mining, transportation, communications and utilities and services. These ratios, representing the growth in taxable Minnesota sales relative to total US sales, are calculated for individual industries and then weighted based on the Minnesota tax data. The weighting has its most significant impact in retail sales; in particular, clothing and motor vehicles have very low weights because the Minnesota sales and use tax does not apply to most purchases in these categories. In other sectors, weighting does not make much difference, and those results are not reported separately.

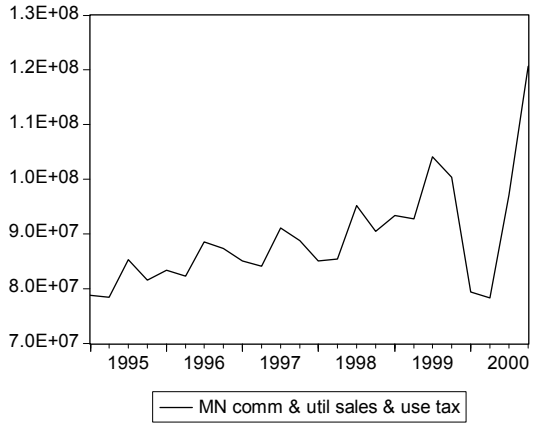
These ratios are then compared with the growth in relevant Minnesota income variables. For retail trade, the comparative measure used is the growth in Minnesota personal income relative to US personal income. For most other sectors, the ratio of

Minnesota to US employment is used. For the agriculture sector, it is the ratio of Minnesota farm income to US farm income.

The method works fairly well in all sectors except manufacturing, where the problem is that the proportion of manufacturing tax revenues accounted for by SIC 39 over the period from 1995 to 2000 grew far too rapidly, rising at approximately 16% per year. It appears that as the years progressed more and more firms in Minnesota were classified as “miscellaneous” instead of in their appropriate industries. Hence we recalculated the ratios without SIC 39, and that yielded reasonable results.

The only sector not represented in this analysis is communications and utilities, where Minnesota sales and use tax receipts show an unusual pattern, as seen in figure V-1. The dip in 1999 and subsequent recovery in 2000 are not tied to any economic variables. This may be an anomaly of the payment and/or reporting process.

Figure V-1



In figure V-2 (next page) the first data column represents the growth from 1995 to 2000 in the ratio of Minnesota sales and use tax receipts to the corresponding measure of national sales or output. The precise variable used for the denominator for each sector is shown in the last column of this table.

The second data column represents the growth from 1995 to 2000 in the ratio of Minnesota employment to US employment, except for retail trade. The relative growth in personal income is used here instead. The third data column represents the tax gap in percentage terms. We would expect this amount to be negative for all industries, signifying a tax gap.

For mining, the numbers for output and employment are small relative to national totals, so we assume the estimated +1.1% gap is not significantly different from zero. For services, Minnesota reported substantial tax receipts in SIC 90 and 99, for which national data are not available. To the extent that these figures should have been included in other SIC classifications some inaccuracy is introduced. Therefore, in this case we also assume that the small +1.3% gap is not significantly different from zero.

The gap in construction appears large, but the amount of tax avoidance is not much different from the amount calculated from the regression equation. The large apparent gap in financial services probably reflects the situation that when profits are rising rapidly, output originating does not track actual sales very closely. We assume in financial services that this is not a gap at all; and it could only account for about 1.5% of total sales and use tax receipts.

Figure V-2

Change in Tax Gap by Major Sector 1995- 2000: Regression Approach

Sector	Relative Growth in Receipts & Sales (%)	Relative Growth in Employment (%)	Relative Growth in Tax Gap (%)	Variable Used to Measure Receipts & Sales
Agriculture	38.5	41.2 c	-2.7	Farm income
Mining	13.3	12.2	1.1 e	Output originating
Construction	-2.3	9.2	-11.5	Construction put in place
Manufacturing	1.1 a	3.9	-2.8	Shipments
Transportation	4.1	4.2	-0.1	Output originating
Wholesale Trade	-5.3	1.7	-3.6	Sales
Retail Trade	1.9 b	3.7 d	-1.8	Sales
Financial Services	-2.9	5.4	-8.1	Output originating
Bus/Pers Services	-6.7	-5.4	-1.3	Output originating
Other Services	4.0	2.7	1.3 f	Output originating

(a) excluding SIC 39. See text
(b) weighted average. See text
(c) average of 1995 and 1996 values
(d) growth in MN/US personal income
(e) figures too small to be significant
(f) MN data includes some taxes in SIC 90 and 99, no national data available

C. METHOD #2: EXAMINING E-SALES

In Minnesota, total E-sales at the retail level in 2000 are $29 billion, and total E-sales at the service sector level are $37 billion. Minnesota personal income in 2000 is just about 2% (1.98%) of total US personal income, so we assume that Minnesota retail sales are also about 2% of total retail sales. Of the $29 billion in E-sales, approximately $5 billion are motor vehicles, which are not included in the Minnesota sales and use tax estimates. Roughly $2 billion are in clothing sales, which is not taxed in MN. That leaves $22 billion of sales in taxable categories, or an estimated $440 million in Minnesota sales. Assuming the 6.5% sales tax rate applies to these purchases, the loss of retail sales and use tax revenues from this source would be $29 million. This figure is then divided by 0.80 to account for other forms of tax avoidance.

A detailed breakdown of E-sales by service sector is not available, but a high proportion of these transactions include booking airline travel, purchasing on-line information services and trading securities. Based on partial information, about two-thirds of service sector E-sales are in these categories plus transportation (note that some firms classified as services businesses sell taxable goods also). This leaves about $12 billion in sales categories that are otherwise taxable in Minnesota. On this basis, Minnesota taxable sales would be $240 million, and lost revenues at a 6.5% tax rate would be about $16 million. This figure is also divided by 0.8 to generate an estimated $20 million lost in sales and use tax revenues in the service sector due to E-commerce. Note that this tax gap estimate is consistent with the Input-Output methodology's results.

D. METHOD #3: REGRESSION EQUATION ESTIMATES

This section describes the underlying regression estimates that are used to estimate tax evasion by detailed 2-digit and 3-digit (SIC code) industries. The underlying structure of the equations used to test for increased tax avoidance can be described as follows. If all the relevant data were available, including Minnesota consumption by detailed SIC classification, it is a simple matter to see how the tax share has changed in recent years. Hence, how much has the degree of tax avoidance increased? Of course, those data are not available; otherwise the study would be much simpler. Instead, taxable sales data are available, but that sidesteps the very problem we wish to investigate.

At the national level, we assume that the figures for retail sales, wholesale sales, manufacturing shipments, and output originating in the rest of the economy are correct and inclusive in the sense that they include both sales on which taxes have not been paid and those that have been paid. Consider, for example, retail sales at building materials and hardware stores (SIC 52). Some of these Minnesota sales are taxed, some of them are exempt (lumber for building new homes, for example) and some escape sales tax.

Assume for the moment that the rate of tax avoidance in Minnesota on sales in SIC 52 had not changed over the 1995-2000 period. In that case, the ratio of Minnesota sales and use taxes to actual Minnesota sales would also be unchanged. The ratio of Minnesota sales to US sales for that category would then depend on several different types of economic variables, including the MN/US ratio for housing permits, personal income, and the unemployment rate. In certain industries, although not building materials, the ratio of stock prices to personal income might also be important because stock prices had a somewhat greater impact on Minnesota income and consumption than was true for the national average. Seasonal factors are also likely to be different for Minnesota than the US, especially for those industries involving construction.

A regression equation is calculated for the ratio of Minnesota tax receipts, assumed to be proportional to Minnesota taxable sales, to total US sales for each category. This ratio is then regressed against the economic ratios mentioned in the previous paragraph. In addition, a variable is included for the ratio of retail E-sales to total retail sales, and the equations are also tested for presence of an additional time trend, although it was usually not significant. The time trend could represent increased tax avoidance for other reasons, or changes in Minnesota tastes and preferences relative to national levels. For that reason, it could be either negative or positive.

Ideally, it would be useful to have detailed quarterly time series of E-sales by 3-digit SIC code level for retail (and wholesale) sales. No such data are available for several reasons. First, quarterly data for retail E-sales start only in 1999-Q4. Although annual data are available for 1998 and 1999, earlier data are estimates. Second, even where quarterly data are available, most of the E-sales are lumped into the "electronic non-store" category without designating what types of goods were purchased. Hence for purposes of this study, it turns out to be more reasonable to use total retail E-sales in each of the retail sales equations rather than trying to estimate individual time series.

The actual regression estimates, together with the parameter values and t-ratios for each term, and the adjusted R-square and Durbin-Watson statistic, are included as part of this report. In many cases, the coefficients for the E-sales term fall below the usual criterion of significance as measured by a t-ratio of 2.0 or greater. The regression results are already substantially higher than the estimates obtained using Methods #1 and #2, so the attempt was made to generate realistic coefficients even if the t-ratios were not as high as usual.

Figure V-3

Alternative Estimate of Change in Sales and Use Tax Gap: 1995-2000

Amounts in Millions of dollars		Estimates of Gap			
Industry	Actual Total Tax	Ratio Method	E-Sales method	Regression method	Preferred Estimate
Agriculture	25	0		0	**0**
Mining	22	0		0	**0**
Construction	55	6		4	**5**
Manufacturing	318	9		19	**12**
Transportation*	403	0		0	**0**
Wholesale Trade	343	12		31	**16**
Retail Trade	1,822	32	36	59	**36**
Financial Services	56	0		6	**0**
Bus/Pers Services	557	7	20	39	**20**
Other Services	177	0		6	**0**
Total	**$3,778**	**$66**	**$56**	**$164**	**$89**

* Tax receipts include communications and utilities, but the ratio method was not employed for those industries.

The results vary substantially. In particular, the results generated by the regression method indicate almost twice as high an estimate as the preferred method. The E-sales method is only a partial total because it does not include any contributions from manufacturing or wholesale trade. The "preferred" estimates take into consideration all

three methods, plus the reliability of the results from some of the individual regression equations.

We have retained the regression equations and included them with this report because we think they show the relative industries where tax avoidance is likely to be highest. We also suggest where Minnesota might concentrate its effort to close this tax gap. It seems likely, for example, that tax avoidance is relatively high in small companies offering personal, business and repair services, even though the ratio figures do not indicate a very large discrepancy. It is quite possible that "off-the-books" employees might not be reflected in the employment totals either. The same general comment applies to small retail and wholesale trade establishments, the manufacturing sector and the printing industry.

Excel Spreadsheets in the appendix give detailed calculations of the tax gap by alternative methods for each industry and give all the regression equations.

E. BREAKDOWN BETWEEN E-SALES AND OTHER TAX AVOIDANCE

AEG also calculated the amount of tax avoidance due to E-sales and other factors for the sample and forecast periods. These results were based on regression estimates that included the E-sales terms, a trend factor and other variables. Initially we tested for the increase in Canadian sales by including the ratio of the Canadian dollar to the U.S. dollar, but that variable was rarely significant. We tested cross-border sales by using the relative price of gasoline believing that consumers would be more likely to fill up in Wisconsin when gasoline prices were low. That variable was not significant either. Finally, the amount of non-store sales, excluding E-sales, was included as an additional variable to measure tax avoidance, but that variable was also statistically insignificant.

The trend factor reflects that the ratio of Minnesota sales taxes to national sales in certain categories fell relative to MN to national income or other economic variables in certain categories, even after taking into consideration the impact of E-sales. It is highly likely that at least part of this trend represents increased tax avoidance. One can speculate on the reason for that increase. It is possible that the sharp cutback in the number of audited tax returns by the IRS over the past decade, especially for small businesses, has also encouraged a number of business enterprises to avoid paying taxes at the state level. It is also possible that the widespread proliferation of on-line sales and the belief that "everyone" is avoiding taxes in this manner has also reduced the payment of other sales and use taxes.

The results in this section are based on (a) pooled cross-section/time-series estimates for all 1-digit categories where tax avoidance was found to be significant and (b) statistically significant trends in some of the 2-digit and 3-digit equations. The results for both methods were approximately the same, with "other" tax-avoidance approximately 20% of total avoidance except for manufacturing where there was no additional reduction noted, and services, where the "other" avoidance was more than 50%. It seems likely that small personal services establishments—such as personal care facilities, small repair shops and similar enterprises—are increasingly avoiding sales taxes on cash transactions. This would also tie in with greater tax evasion at the Federal level. If these small firms are underreporting the total amount of transactions,

they would also be quite likely to underreport the amount of goods or services they bought and sold in order for all figures to appear consistent.

Figure V-4

E-Sales and Other Tax Avoidance by Sector: Change 1995 - 2000			
	Tax Avoidance due to		
	E-sales only	Other	Total
Construction	4	1	5
Manufacturing	12	0	12
Wholesale Trade	11	5	16
Retail Trade	30	6	36
Services	9	11	20
Total	**$66**	**$23**	**$89**

F. PROJECTIONS THROUGH 2007

The regression equations are also used to calculate the increase in the tax gap from 2001 through 2007. The detailed quarterly estimates for each equation are given on the accompanying spreadsheet, together with the baseline projections. These results are summarized in figure V-5.

To calculate the projections, a baseline projection of tax receipts is estimated for each equation based on our forecasts of national trends in the overall economy and various industries. For purposes of this simulation, it is assumed that the corresponding variables for Minnesota grew at the same rate as their US counterparts. In these calculations, it is also assumed that E-sales relationships remained unchanged.

Second, the equations are resolved with estimated increases in E-sales, and the difference is calculated. The calculated gaps are then scaled down by the same proportion as is found in the 1995-2000 period, with the ratios being the preferred estimates relative to the sum of the regression estimates. In general, these ratios are close to 0.5 for all sectors except construction, where a ratio of 1.0 is used. The quarterly figures are then summed or averaged to obtain the figures shown in figure V-5.

Figure V-5

Projected Change in Sales and Use Tax Gap: 2000 - 2007 (millions of dollars)

	Retail		Construction		Wholesale		Manufacturing		Services		Total (2)	
Year	Gap	%	Gap	%	Gap	%	Gap	%	Gap	%	Gap	%
2000 Baseline												
2001	$5.8	0.4%	$0.3	0.5%	$1.0	0.4%	$0.2	0.1%	$3.8	0.8%	**$11.1**	0.3%
2002	$13.9	1.0%	$0.7	1.2%	$2.8	0.9%	$0.7	0.4%	$8.2	1.6%	**$26.3**	0.6%
2003	$21.2	1.5%	$1.0	1.8%	$4.1	1.3%	$1.5	0.8%	$12.8	2.4%	**$40.6**	0.9%
2004	$29.7	2.0%	$1.4	2.5%	$5.5	1.6%	$2.5	1.3%	$18.2	3.3%	**$57.3**	1.3%
2005	$39.3	2.5%	$1.8	3.2%	$7.0	1.9%	$3.4	1.7%	$24.3	4.2%	**$75.8**	1.6%
2006	$50.5	3.0%	$2.2	3.9%	$8.7	2.2%	$4.5	2.0%	$31.4	5.1%	**$97.3**	1.9%
2007	$63.2	3.6%	$2.6	4.7%	$10.4	2.5%	$5.7	2.4%	$39.5	6.1%	**$121.4**	2.2%

(1) "%" is gap as a percent of total baseline tax estimate
(2) Total includes categories not shown separately (no estimated tax gap)

The estimated growth in E-Sales is summarized in Figure V-6. These data are based on a modified logistics curve that also takes into consideration the slowdown in the economy during the 2001 recession. That effect is seen mainly in the manufacturing sector, since retail sales slowed far less than usual during that recession.

Figure V-6

Estimated and Projected E-sales by Sector for the United States (billions of dollars)

	Retail	%	Services	%	Wholesale	%	Mfg	%
1995					10.0	0	0.0	
1996					26.0	160%	118.0	
1997	$1.5		$5.0		60.0	131%	355.0	201%
1998	$5.0	233%	$15.3	206%	109.0	82%	593.0	67%
1999	$15.0	200%	$25.3	65%	183.0	68%	730.0	23%
2000	$28.9	93%	$37.3	47%	213.0	16%	777.0	6%
2001	$35.9	24%	$44.8	20%	239.0	12%	808.0	4%
2002	$43.8	22%	$52.8	18%	267.0	12%	857.0	6%
2003	$52.6	20%	$61.3	16%	297.0	11%	925.0	8%
2004	$62.0	18%	$70.5	15%	326.0	10%	1008.0	9%
2005	$72.6	17%	$80.3	14%	356.0	9%	1099.0	9%
2006	$84.2	16%	$90.8	13%	384.0	8%	1187.0	8%
2007	$96.8	15%	$101.7	12%	411.0	7%	1282.0	8%

1. Actual Census data are available for 1998-2001 and the first half of 2002 for retail; 1998-2000 for services; 1998-2000 for wholesale; and 1999-2000 for manufacturing. All other figures are estimated.

2. Predictions are based on modified S-curves for all sectors. The dip and recovery in the % change for manufacturing reflects the recession and recovery.

The 22% gain in 2002 for retail sales represents the actual increase for the first half of this year. It is assumed that the second half gain will be the same percentage. This figure is far lower than the estimates used by Donald Bruce and William F. Fox in their paper[6]. They state that the figures they used are derived from various estimates prepared by Forrester Associates.

The figures quote in the Bruce and Fox paper show increases of 63% for business transactions and 57% for consumer transactions for 2002. Census data for business transactions are not available, but it can immediately be seen that the 57% figure for consumer transactions is far away from the actual increase of 22% so far this year. Similarly, the projected gains of 56% for business transactions and 44% for consumer transactions in 2003 appear to be similarly high. In AEG's view these figures are not worthy of serious consideration.

Figure V-7 shows the projected tax avoidance from E-Sales and other factors for 2007. The methodology used to calculate this split is the same as was used for the 1995-2000 period. Again note that these increases in tax avoidance are in addition to the tax gap that already exists in 2000.

Figure V-7

E-Sales and Other Tax Avoidance by Sector: Change 2000 - 2007

	Tax Avoidance due to		
	E-sales only	Other	Total
Construction	2	1	3
Manufacturing	6	0	6
Wholesale Trade	8	2	10
Retail Trade	51	12	63
Services	27	12	39
Total	**$94**	**$27**	**$121**

6 Op. Cit.

VI Technical Appendix: Constructing and Maintaining The Microsimulation Database

Part of the Minnesota Sales and Use Tax Gap Project involved the development of two databases (Level II and Level III). This document presents a description of the database development process for both Level II and Level III databases. The only difference between the Level II and Level III databases is that the Level III database provides information for both filing and non-filing taxpayers. There is also a Level I database, but that is an extraction of transaction data directly from DOR files and does not require extensive instructions.

Level II Database

Level II database development steps include:

- Create an audit sample database through DOR record matching.
- Determine the statistical reliability of audit sample information.
- Relate the audit sample information to population of sales and use tax returns for 2000.

Constructing the Level II database involves detailed analytic work to statistically merge audit information with larger transaction files constituting the Level I database. The approach is presented below in addition to the methodology for obtaining estimates of the sample errors and levels of confidence.

The methodology starts with the determination of the stratum boundaries relating to the tax classes. This is necessary to line up the two databases (audit and transaction) by the same tax classes. The second step involves determining industry groups, necessary to line up the Level I and Level II databases according to industry. Ideally, industry groups should be reasonably narrow, but this is constrained by the size of the audit sample by industry. The third step randomly allocates audit sample records by strata to tax returns conditional upon a sufficient number of audit records from which to draw. Finally, the Neyman Allocation sampling methodology is used to determine the sampling error of the audit sample by strata, given that this sample, itself, is predetermined.

Create Audit Sample Database through DOR Record Matching

The first step in the Level II database development effort is to create the audit sample database through DOR record matching. Specifically, a matched file is created using the following DOR databases:

- Audit transaction records for 1999, 2000 and 2001

- Revenue audit ledger records for 1997, 1998, 1999, 2000, 2001 and 2002
- Inquiry data for 1997, 1999 and 2000
- Transaction data for 1995, 1996 and 1998

Databases are matched using taxpayer primary ID as a matching key. Following is a discussion of the information contained in each of these databases as well as the methodology for developing the matched database.

Audit Transaction Records for 1999 2000, and 2001 The audit transaction records provide detailed information relating to audits and refund claims. Specifically, the "field audit" information forms the base database used as a starting point for creating the matched database. This file contains detailed audit information including whether there was an audit assessment for sales or use tax, interest, and penalties, but does not contain information relating to the period covered by the audit. In addition, the file contains information on refund claims and whether the claims were allowed by DOR.

The field audit information currently is part of a larger database containing data on refund claims and other types of audits. It is maintained by DOR in Excel format. The field audit data is in separate sheets by state region for a given year. The field audit data were extracted from each sheet by region and combined with the field audit data for each of the three years. This process produced 4,492 separate field audits during this three-year period.

Revenue Audit Ledger Records for 1997, 1998, 1999, 2000, 2001 and 2002 The revenue audit ledger records data provided by DOR contain information on payments relating to audits as well as the period covered by the audit. There were 1,685 records for 2002, 2,244 records for 2001, 2,054 records for 2000, 1,608 records for 1999, 1,832 records for 1998 and 1,984 records for 1997. These data files contain only four fields: 1) the primary ID, 2) the beginning period covered by the audit, 3) the ending period covered by the audit and 4) the audit payment.

Revenue audit ledger records were matched to audit transaction records using taxpayer primary ID as a matching key. The resulting matched file contained 3,714 records providing detailed audit assessment information, the period covered by the audit and the actual payments relating to the audit. It is important to note that when the field audit did not result in any assessment, the audit transaction record was retained in the database with no match from the revenue audit ledger. Obviously, if there was no audit assessment, there would be no payment associated with the audit. The retention of these field audits with no assessment was essential to preserve the probability of an audit resulting in no assessment for the final audit sample database.

Inquiry Data for 1997, 1999, and 2000 and Transaction Data for 1995, 1996 and 1998 The next step in developing the matched database is to match the

matched audit file with the actual tax return. This is necessary to establish relationships between the audit assessment, audit payments, and the fields on the actual tax return, such as gross sales and sales and use tax. The result of matching the audit assessment and payment information with the tax return yielded the final audit sample database.

The period covered by the audit determined the year to which the matched audit record would be linked to the return. For example, if the audit covered a three-year period such as 1998, 1999, and 2000, the matched audit record would be matched with the 2000 tax return— the most recent tax return available. Similarly, if the audit covered the two-year period such of 1997–1998, the matched audit record would be linked with the 1998 tax return. This matching process is applied to every year's tax return database. For this project, there were 3,272 matched records and 442 non-matches. The audit sample database of 3,272 records forms basis for the development of the Level II database.

Determine Statistical Reliability of Audit Sample

The next step in the Level II database development process is to determine the statistical reliability of the audit sample. The Neyman Allocation statistical sampling methodology is used to evaluate the reliability of the Audit sample. This methodology minimizes the sample size required for a given level of sampling error. The steps involved in performing this evaluation include:

- Stratify the audit sample records according to tax.
- Create sampling equations for each stratum using the Neyman Allocation method.
- Vary the level of error used in the Neyman Allocation equations until its results match the Audit sample results.

The audit sample was not based upon a stratified random sampling methodology, and assessing the statistical reliability of the sample using the Neyman Allocation method essentially assumes a "reallocation." The use of this technique is to provide general benchmarks for the statistical reliability of the sample. The Neyman Allocation method is appropriate if a smaller overall sample size is desired to achieve a specified error level for the sample. This could be very valuable in targeting audits to specific industries where the number of audits would be limited. AEG developed and provided DOR with software that allows the user to specify different error levels to determine the optimal sample by industry and tax class.

Relate Audit Sample Database to Population of Sales and Use Tax Returns for 2000

The final step in developing the Level II database is to relate the audit sample database to the population of sales and use tax returns for 2000. The steps involved include:

- Determine tax class stratum boundaries.
- Determine industry groups.
- Randomly allocate Audit sample records by strata to tax returns.

Determine Tax Class Stratum Boundaries The tax strata were developed using statistical techniques to determine the appropriate stratum boundaries. Similar to the grouping of industries, there is the same constraint with regard to the number of stratum boundaries. Specifically, the stratum boundaries were determined largely with the consideration of the size of the audit sample. Expanding the number of stratum boundaries beyond the number used resulted in a low number of audit records in the higher strata.

The method used to determine the stratum boundaries for the tax classes is the *cumulative of the square root of the frequency* method. There are four steps involved in the determination of stratum boundaries:

- Stratify the returns within the population of items to be sampled ($0 – $1,000) using a set dollar interval between the strata ($1,000). We use $1,000 for the current example to limit the number of strata that need to be displayed.

- Determine the frequency $f(y)$ for each dollar range. This is nothing more than the number of returns within each dollar range.

- Calculate the square root of the frequency for the first sampled range ($0 - $1,000). Next, calculate the square root for the next sampled range ($1,001 - $2,000). For the higher strata, the results were adjusted by computing the square root of the interval distance in order to compute the cumulative square root of the frequency. The results are accumulated for this column by adding each result to the preceding number. This process is continued for each of the sampled strata.

- Once the cumulative square root of the frequency for the last sampled strata is calculated, each stratum can be determined. The strata are determined by computing equal distance in the cumulative square root of the frequency.

Determine Industry Groups The original data for the population and the audit sample were not modified in any way. Both of these data sources contain four-

digit SIC codes. The groupings by industry were for stratification purposes. The industry groups were determined largely by the size of the audit sample for a particular industry group. For example, retail trade is not audited as intensively as other industries. As a result, there were not many audit records when that industry group was broken into more detail such as the two-digit Retail Trade categories.

Determining the industry group for the Level II database development is constrained by the existing audit sample. The audit sample consists of 3,272 records. In order to provide for a sufficient number of returns for each stratum from the audit sample, it was necessary to combine industries to preserve statistical reliability of the results. Generally, the combination of industries was determined by ensuring that the number of returns within a given stratum was 5 or greater. This process resulted in 14 industry groups. Some industries were not well represented in the audit sample, including Retail Trade. In this case, it was necessary to combine a number of industries that contained a large number of returns in the population.

Randomly allocate Audit Sample Records by Strata to Tax Returns For companies having both tax return and audit data, perform an exact match using primary ID as the match key. For companies having no audit data, the following process is employed:

- Stratify the target tax return using the stratification criteria.
- Locate the corresponding stratum in the audit sample database.
- If the stratum in the audit sample database contained fewer than 10 returns, include information from the tax class above and below the stratum (keep adding strata) until the audit sample information contained at least 10 returns.
- Generate a random number between 0.0 and 1.0.
- Relate the random number to the numerical location of an audit sample record in the appropriate stratum (1 to 10 for a stratum containing 10 returns).
- Assign the audit sample information to the target tax return.

In order to develop the Level II database, it was necessary to attach Audit sample records to the population returns for 2000. It is useful to note that each stratum stands alone and the allocation procedure associates taxpayer and audit records for a particular stratum. The probability of selection of an audit record within a stratum is based on the number of audit records within that stratum. A uniform random distribution between 0 and 1 was divided into equal intervals according to the number of audit records in the stratum. For example, if the stratum had 20 audit records, the 20 intervals ran 0–.049999, .05–.9999, etc. Each interval was assigned to a unique audit record. This gives all audit records an equal chance of being picked when a random number between 0 and 1 is drawn. Successive draws while stepping through taxpayer records pick an audit

record and assign it to the current taxpayer record. The audit results are then appropriately ratioed to the taxpayer record based upon the relationship of the audit information to gross sales contained in the audit sample.

The strata are designed to be homogeneous for the group of audit records (and taxpayers) in each stratum. Therefore a random association of audit results and taxpayers is appropriate, but only within a stratum. This method would not work for the population alone, for example. Homogeneous strata must first be established.

The resulting database is the final Level II database, which can be used to generate an estimate of the amount of sales and use tax that would be collected if every taxpayer were audited. This is accomplished by calculating the relationships between the audit payment and audit assessment information to gross sales contained in the audit sample record. These relationships are computed for each tax return for 2000. The computation of these relationships provides an estimate of the amount of sales and use tax that would be collected if every taxpayer were audited in 2000.

For the tax gap project, audit results from the audit file were assigned to taxpayer records as described above. An audit result for any given taxpayer is an estimate of what the audit results would be if taxpayers similar to the taxpayer in question were audited. While specific taxpayers in the file have an expected auditing result assigned to them, the results do not speak to what would happed if that particular taxpayer were audited, only to aggregations of similar taxpayers. Think of this like an actuarial process; for a group of 1,000 men aged 50. We may know on average they will live to 78 (or whatever) but not know how long any specific one will live. Yet we may write an expected age of 78 in a record for each of them.

Level III Database

Level III database development includes:

- Develop an Input-Output model of Minnesota sales and use the tax system to estimate total Minnesota sales and use tax that should be generated.
- Reduce the total sales and use tax estimate by amount of tax collected—total tax gap estimate.
- Reduce the total tax gap estimate by tax gap estimate generated using the Level II database.
- Divide the remaining (uncollected) tax into taxes owed by individuals and businesses.

Develop Input-Output Model of Minnesota Sales and Use Tax System

The steps involved in the development of the Input-Output model of the Minnesota Sales and Use Tax system include: 1) database development and extrapolation, 2) tax law parameter development and 3) tax calculator development.

Database Development and Extrapolation

The US Bureau of Economic Analysis' (BEA) Input-Output (IO) account data for the US economy for 1992 form the basis for the underlying database. These data include:

- The Intermediate Business Purchases (Use) matrix, which is a matrix containing 491 rows and columns, where the rows represent intermediate business sales of goods and services and the columns represent industry purchases of intermediate business goods and services.

- The Business Capital Flow Table for New Structures and Durable Equipment matrix that contains 163 categories (columns) of new structures and durable equipment covering 64 industries (rows), which purchase new structures and durable equipment.

- The Personal Consumption Expenditures (PCE) is a single column of consumer spending on 85 major products and services with further detailed breakdowns by SIC/NAICS codes.

In order to provide analysis capabilities for years beyond 1992, these databases are "aged" or extrapolated to 2000 based upon detailed national industry growth from 1992 to 2000. After these databases are aged to 2000 levels of economic activity for the nation, they are benchmarked to Minnesota levels.

Different methods are employed to benchmark these separate US databases to Minnesota levels. This process is discussed below.

Intermediate Business Purchases: Minnesota Levels The Regional Purchase Coefficients (RPC) method was used to scale the US Intermediate Business Purchases matrix to current Minnesota levels of economic activity. First, RPCs are computed for each of the 491 industries. The RPC is the ratio of Minnesota to US employment for a particular industry. County Business Patterns data on employment for 2000 for Minnesota and the nation by detailed SIC categories was mapped into the 491 I-O industries. Industry-specific RPCs (ratios of Minnesota to US employment for 2000) are used to benchmark the intermediate matrix columns to Minnesota levels.

The columns of the matrix represent industry purchases for Minnesota, whereas the rows of the matrix represent intermediate sales of goods and services. The resulting columns represent the 491 industries' purchases of intermediate goods and services for Minnesota. Summing up the 491 entries for a particular row provides an estimate for Minnesota sales for that intermediate good or service. The resulting 491 rows represent Minnesota intermediate business sales of goods and services, which are the raw materials, semi-finished products and services sold to Minnesota businesses. Thus, state business intermediate purchases are obtained when the columns are benchmarked to Minnesota economic activity. The resulting rows provide estimates of sales of these intermediate goods and services.

Business Purchases of New Structures and Durable Equipment – Minnesota Levels The method used to scale the US Capital Flow Table (CFT) matrix to Minnesota levels is essentially the same method used to benchmark the Intermediate Business Purchases matrix to current Minnesota levels. The RPCs by industry are the ratios of Minnesota employment to US employment. These RPCs are computed for each of the 491 industries contained in the CFT matrix. Multiply these ratios by the US levels contained in the matrix to obtain estimates of Minnesota purchases of new structures and durable equipment. The end result of this process provides estimates of purchases of 163 categories of new structures and durable equipment for Minnesota's businesses for 64 industries.

Personal Consumption Expenditures: Minnesota Levels Scaling the US personal consumption expenditures to Minnesota levels requires two steps. For the tax gap project, the 1992 US consumption categories were grown to 2000 levels based upon national industry GDP growth by specific category. For example, retail trade and service categories were grown based upon the respective growth in GDP for these industries. Next, each of the US 85 major consumption estimates were scaled to Minnesota levels by multiplying each by the ratio of Minnesota population to US population for 2000.

For the 85 major categories of consumption expenditures, it is possible to generate further breakdowns of these major categories into minor categories thereby expanding the number of separate categories for consumption if necessary.

Tax Law Parameter Development

There are three different sets of tax law parameters for each of the three separate databases. The three separate databases are: 1) business intermediate purchases with 491 categories of goods and services; 2) business purchases of producers durable equipment with 163 major categories; and 3) personal consumption expenditures with 85 major categories of goods and services. In addition, each set of tax law parameters for each of these three databases has two components. The first component is the statutory tax rate for a particular

good or service. The second component is the “percent in base” for a particular good or service.

Business Intermediate Purchases The business intermediate matrix contains 491 rows that represent sales of intermediate goods and services to Minnesota businesses. The columns represent the industries making those purchases. The first component of the set of tax law parameters for business intermediate sales applies statutory tax rates to each of the 491 rows based upon whether a category is subject to the Minnesota Sales and Use Tax. Thus, there are 491 tax law parameters that reflect the statutory rate for each category.

The second component of the set of tax law parameters is the “percent in base.” This set of tax law parameters consists of a 491-by-491 matrix of elements. In practice, this second component matrix collapses to only 491 numbers. An exception occurs when a particular industry is exempted from purchases of specific goods or services that are subject to tax in other industries. An example of this could be purchases by businesses in the agriculture industry that are provided exemptions for certain goods that are subject to tax for businesses in other industries.

The values for the entries for this second set of tax law parameters usually take on values of zero or one. Given the level of detail for these 491 separate categories, the sales and use tax treatment for a particular category is that it is either subject to tax or exempt. There are usually few exceptions where a particular category contains several goods or services and the tax treatment may exempt some and tax others within this category. In such a case, an estimate iss made of the percent of that category subject to tax. In this case, the value would be between zero and one.

Business Purchases of New Structures and Durable Equipment The matrix for purchases of durable equipment contains 64 rows that represent purchases of durable equipment for Minnesota businesses. The columns represent the specific categories of durable equipment. The first component of the set of tax law parameters for sales of durable equipment applies statutory tax rates to each of the 163 columns based upon whether a category is subject to sales and use tax. Thus, there are 163 tax law parameters that reflect the statutory rate for each category.

As with intermediate sales, the second component of the set of tax law parameters is the “percent in base.” This set of tax law parameters consists of a 64-by-163 matrix of elements. In practice, this second component matrix collapses to 163 numbers. Exceptions occur when a particular industry is exempted from purchases of a specific category of durable equipment that is subject to tax in other industries. An example of this could be businesses in the manufacturing industry that are provided sales and use tax exemptions for

purchases of certain durable goods where these same durable goods are subject to sales and use tax for businesses in other industries.

Personal Consumption Expenditures The column for personal consumption expenditures contains 85 separate categories. The first component of the set of tax law parameters for personal consumption expenditures applies statutory tax rates to each of the 85 categories for goods and services based upon whether a category is subject to the sales and use tax. Thus, there are 85 tax law parameters that reflect the statutory rate for each category.

The second component of the set of tax law parameters is the "percent in base." This set of tax law parameters consists of 85 elements. The values for the entries for this second set of tax law parameters usually take on values of zero or one. There are usually a few exceptions where a particular consumption category contains several goods or services and the tax treatment may exempt some and tax others within this category. When this occurs, make an estimate of the percent of that category subject to tax. In this case, the value for the "percent in base" was between zero and one and the statutory tax rate was set at the rate for those items subject to tax.

Tax Calculator Development

Basically, the sales and use tax calculator consists of applying the tax law parameters to each database and generating output tables that present the results of the computations. This calculator generates an estimate for the total amount of sales and use tax that should be generated by the Minnesota sales and use tax system.

Reduce Total Sales and Use Tax Estimate by Amount of Tax Collected: Total Tax Gap Estimate

The total tax gap is estimated by subtracting the amount of sales and use tax that is collected from the total that should be generated from the sales and use tax system using the I-O model.

Reduce Total Tax Gap Estimate by Tax Gap Estimate Generated Using the Level II Database

The tax gap estimates generated using the Level II database represents the tax gap for current filers. The estimates obtained by subtracting the Level II tax gap from the total tax gap represents the tax gap attributable to non-filers. This non-filer tax gap is referred to as the Level III tax gap.

Divide Remaining (Uncollected) Tax into Taxes Owed by Individuals and Businesses

In order to develop the Level III database, it is necessary to obtain estimates of the breakdown of the Level III tax gap between consumers and businesses. For households, it is presumed that the Level III tax gap is attributable exclusively to retail trade associated with E-commerce and catalog sales. For businesses, the Level III tax gap is attributable to all industries. Thus, the Level III tax gap associated with E-commerce and catalog sales represented only a portion (although significant) of the total Level III tax gap for the Retail Trade.

For this project, the estimate of the Level III tax gap attributable to E-commerce and catalog sales was developed using the US Census Bureau's recent survey of E-commerce and catalog sales for retail trade. The national results from this survey were scaled to Minnesota levels based upon the ratio of Minnesota's retail trade sales relative to the nation for 17 categories of goods. The tax treatment of the 17 categories was then determined by assessing the percent in base for a particular category. Finally, estimates of the split between consumers and business were obtained through reliance of the relative breakdown for comparable categories from the Input-Output results.

The results of this analysis (presented in table III-3 in the main text) indicate that of the retail trade estimate of $43,365,375 that 70.5% or $30,799,564 is attributable to E-commerce and catalog sales. The estimate of the consumer portion of this is $16,496,217 whereas the business portion is $14,303,348. Thus, the consumer portion of the Level III tax gap attributable to E-commerce and catalog sales is 53.6%, and the business portion is 46.4%.

Level III Database Development

The preceding analysis that yielded the Input-Output estimates and E-commerce-catalog sales estimates provides the targets for the development of the Level III database. The development of the Level III database is naturally broken down into two components: households and businesses.

Level III Household Database The Level III household database was created by constructing joint distributions of the number of annual transaction by a household, the type of activity (E-commerce and/or catalog purchases), the amount of each purchase, and the number of purchases in a transaction. A household record was created by picking simultaneously from each distribution.

Then, by setting the number of records to a manageable size of 60,462, sufficient to represent all the variations found, weights were assigned to reach the calibration targets. The Level III household database is calibrated to an E-commerce target gap of $16,496,323 and catalog purchase target gap of $58,258,433, representing the total use tax attributable to Minnesota households not filing use tax returns for these purchases.

The distribution of purchases is based upon 17 categories of likely purchases from Census and Consumer Expenditure Survey data. These are presented in the table below.

Minnesota Households E-Commerce and Catalog Purchases: 2000

Source	Gross Taxable Sales E-Commerce*	E-Commerce Use Tax Due	Gross Sales Catalog*	Catalog Use Tax Due	Total E-Commerce & Catalog Use Tax Due	Average Item price**	Number of E-Commerce + Catalog Purchases***
Electronic Shopping and Mail-Order Houses							
Books and magazines	11,760,359	712,701	12,234,578	741,425	1,454,126	103.50	231,832
Computer hardware	31,210,095	1,891,393	104,661,625	6,342,576	8,233,969	1,603.26	84,749
Computer software	5,726,387	347,030	12,587,744	762,827	1,109,857	51.75	353,892
Drugs, health aids, and beauty aids	9,208,894	558,077	159,021,075	9,636,801	10,194,878	103.50	1,625,391
Electronics and appliances	23,429,010	1,419,844	51,954,937	3,148,510	4,568,354	300.16	251,151
Food, beer, and wine	4,787,171	290,112	11,276,790	683,382	973,494	385.03	41,722
Furniture and home furnishings	13,663,897	828,059	89,933,613	5,450,047	6,278,106	1,603.26	64,618
Music and videos	31,313,735	1,897,674	77,697,895	4,708,553	6,606,227	103.50	1,053,239
Office equipment and supplies	4,074,074	246,897	15,639,050	947,739	1,194,635	1,603.26	12,296
Toys, hobby goods, and games	18,447,502	1,117,955	50,144,575	3,038,800	4,156,755	103.50	662,717
Other merchandise	51,902,149	3,145,372	329,701,895	19,980,191	23,125,563	103.50	3,686,952
Other nonmerchandise	24,269,366	1,470,771	46,494,166	2,817,583	4,288,354	103.50	683,697
Traditional Retail Trade							
Motor vehicles and parts dealers	7,115,272	431,199	-	-	431,199	776.27	9,166
Electronics and appliance stores	11,987,953	726,493	-	-	726,493	300.16	39,939
Building materials and garden equipment and supplies stores	5,483,567	332,315	-	-	332,315	103.50	52,981
Sporting goods, hobby, book and music stores	9,210,928	558,200	-	-	558,200	103.50	88,993
Miscellaneous store retailers	8,617,383	522,230	-	-	522,230	103.50	83,259
Total	**$272,207,743**	**$16,496,323**	**$961,347,944**	**$58,258,433**	**$74,754,756**	**$136.66**	**9,026,594**

*This is before the $770 exclusion. The tax due is on the lower base after the exclusion.
**Where a similar price is shown for multiple items, it represents the average for those items, not the item's individual price.
***On average 1.61 items are purchased per transaction. Thus, 9.0 million purchases involves 5.6 million transactions.

The process of creating distributions began by determining to probability of selecting each type of merchandise for purchase. For each it is the number of purchases divided by the total number of purchases overall. For example, there was a 2.6% probability of assigning a “books and magazines” purchase (231,832 divided by 9,026,594) whereas there was a 0.9% probability of assigning a “computer hardware” purchase (84,749 divided by 9,026,596).

The next step involved drawing a uniform random number (RN) between zero and one to determine the selection of a particular purchase. This was accomplished by spreading the probability of a purchase to cover a unit interval. For example, the “books and magazines” was selected if the RN was greater than 0 and less than or equal to 2.6%. The “computer hardware” category was selected if the RN was greater than 2.6% and less than of equal to 3.5%.

Once a purchase category was selected for a given household, it was assigned a purchase amount. This was accomplished by picking from a log-normal distribution about the mean value for that category.

The result of this process was a database of 60,462 records weighted to add to total purchases of $1,233,555,687, split between $272,207,743 in E-commerce and $961,347,944 in catalogs. After adjusting for the $770 exclusion, household by household, the tax base of $253,789,581 in E-commerce and $896,283,433 in catalog sales was reached. (Refer to figure IV-1 in the main text for more details.)

The calibration also yielded 5,606,580 total transactions with the average 12.8 transactions per household and 1.61 purchases per transaction, resulting in 438,492 households engaging in E-commerce or catalog sales that would be subject to the use tax.

Level III Business Database The development of the Level III business database used the target of $81,170,662 ($96,666,879 minus $16,496,217 for households) as the basis for the total tax attributable to Minnesota businesses not filing tax returns for their purchases. Each industry had a separate tax target. In the case of the “F.I.R.E.” industry group, the target was set at zero.

The basis for creating non-filing tax returns for the Level III database relied upon the profile of current specific tax returns. It was assumed that these non-filers have profiles similar to the smallest taxpayers. Specifically, the lowest decile for each major industry group was first computed based upon the amount of sales and use tax paid. Tax returns were randomly selected from this lowest decile until the target level of tax for each industry group was achieved. In order to reduce the sample size for the Level III business database, each record was assigned a weight of three, thus representing a one-third sample of the population of non-filing business tax returns. The breakdown between the sales tax and the use tax are derived based upon the results of from the returns

selected for inclusion in the sample. The table below provides the results for the Level III sample of business tax returns.

Level III Business Sales and Use Tax Gap Estimates for 2000 by Industry

Total Sales and Use Tax	Sample Returns	Population Returns	Gross Sales	Sales and Use Tax Gap	Sales Tax Gap	Use Tax Gap
Agriculture	247	741	90,184,539	316,424	90,712	225,713
Mining	29	87	81,965,346	330,214	26,159	304,055
Construction	254	762	492,384,333	2,854,382	635,807	2,218,574
Manufacturing and Wholesale Trade	2,414	7,242	17,932,149,633	29,678,701	12,427,209	17,251,492
Transportation, Utilities	68	204	580,096,842	1,119,072	118,048	1,001,024
Retail Trade	5,427	16,281	4,369,053,461	27,065,113	17,692,872	9,372,242
FIRE	0	0	0	0	0	0
Services	4,439	13,317	3,018,861,892	20,895,391	9,700,238	11,195,154
Total	**12,878**	**38,634**	**26,564,696,046**	**82,259,297**	**40,691,045**	**41,568,254**

VII Comparison with Other E-commerce Estimates: The GAO Report to Congress and The Fox Study

Overview

AEG's approach to measuring sales and Sales and Use Tax losses to E-commerce rests upon estimates of overall economic activity in Minnesota and estimates of E-commerce within that framework of total activity. The results are corroborated with Census' direct estimates of E-commerce, recognizing that the Census data omits a portion of internet sales by companies that are not primarily in the internet sales business.

Also, AEG's regression and ratio approaches measure the change in tax losses from E-commerce and from other sources from 1995 and project the tax loss to 2007. The change in E-commerce from 1995 to 2000 is also a measure of its absolute level in 2000 because there were no significant internet sales in 1995.

AEG uses its carefully constructed matrix of what is taxable and what is not (see discussion in main text) and relies on Level I and Level II databases that estimate compliance rates by industry and by other characteristics. Thus, AEG has accounted for payments of use tax, for example, by Minnesota businesses purchasing from other businesses over the internet. The result is an E-commerce estimate grounded in both the level of economic activity and the law and practice of Minnesota and its taxpayers.

The GAO Study

In its report to Congress[1] GAO examined alternative estimates of all "remote sales" and internet sales. The remote sales embraced most sales to a state from outside its borders, of which internet sales or E-commerce, is a subset. The study examined each state in summary detail, accounting in only a general way for its tax rates, its taxable activity, and its assumed rates of compliance.

It offered a range of low to high tax losses depending upon which source of internet estimates it relied upon (Forrester was the highest) and varying the compliance assumption and other factors. It acknowledged that its range of estimates is wide. For 2000 in Minnesota, GAO estimated a low of $49 million in lost taxes from all remote sales, of which losses from internet sales were $5 million. Its high 2000 estimates were $192 million and $72 million, respectively. AEG's estimate of the internet tax loss in 2000 is $66 million, which falls within the GAO range of $5 million to $72 million.

[1] "Sales Taxes: Electronic Commerce Growth Presents Challenges; Revenue Losses are Uncertain," United States General Accounting Office, June 2000, GAO/GGD/OCE-00-165.

For 2003, GAO's estimated range for Minnesota includes a low tax loss of $129 million, with $19 million from internet sales, to a high tax loss of $489 million, including $232 million related to the internet. AEG's internet tax loss estimate for 2003 of approximately $95 million falls within the GAO range of $19 million to $232 million. Recall that the high-end GAO growth to 2003 relied upon Forrester. GAO said the study by Forrester claiming national internet sales growth from 1999 to 2003 of $1.23 billion to $10.8 billion (ten-fold) was based upon a "limited empirical basis." In commenting upon other estimates, GAO stated that the Fox study[2] assumed that more than half of business-to-business sales are taxable and that compliance is low, conditions that may vary greatly among the states.

The wide ranging estimates produced by GAO attest to the difficulty of calculating tax losses for every state in the country. Both the uncertainty of the overall estimates of national activity and of the specific details of state tax law and regulations contribute to GAO's uncertain estimates.

The Fox Study

The results of the Fox study were largely driven by projections of E-commerce that turned out to be exceptionally high when compared to actual sales after-the-fact. In addition, the Fox study covered 50 states and Washington DC with general assumptions regarding the taxability of broad categories. Sales tax treatment of intermediate business transactions vary significantly across states and is usually highly complex and very specific for narrow categories of goods and services. Thus, a general tax treatment of broad categories across all states, even if accurate on average is unlikely to be accurate for any specific state. Finally, assumptions about what share of taxable sales that could be collected in the absence of E-commerce is highly speculative.

Fox Study Methodology

Key inputs to estimating the tax base loss for E-commerce transactions include: forecasts of E-commerce sales, identification of the sales taxable components of these sales, assumptions about what share of taxable sales could be collected in the absence of E-commerce and estimates of the share of taxes due that can be collected.

E-commerce sales were drawn from Forrester Research Inc.'s annual forecasts for the years 1999 through 2003 for 24 categories of business-to-consumer (B2C) sales and 13 categories of business-to-business (B2B) sales. Forrester projected a rapid compound growth rate of 83.7 % annually through 2003. B2B

[2] See "State and Local Tax Revenue Losses from E-Ecommerce: Updated Estimates," Donald Bruce and William F. Fox, University of Tennessee, Knoxville, October 2001, published on the Web at http://www.statestudies.org/ecomreport.pdf.

sales were expected to dominate E-commerce activity, representing 90.3 % of the 2003 total. The Fox analysis adjusted Forrester's forecasts to net out purchases by businesses and residents in non-sales-taxing states. The assumption was that the share of E-commerce sales in these states was proportionate to their share of the national population. The remaining transactions were assumed to be made by residents and businesses in sales-taxing states.

Sales tax bases differ by state, and Forrester used relatively broad categories that made assumptions about the percentage of sales for each sales category that would be taxable on average across the U.S. For sales that were expected to occur through E-commerce, major exempt purchases on B2C transactions included most leisure travel, including airline tickets purchased through E-commerce, much of the food and beverage purchases (at least 27 states exempt food for consumption at home), some health and beauty expenditures (medical expenditures are exempt in most states) and a portion of apparel (part of apparel expenditures are exempt in some states). Based on the specific assumptions adopted, 70.2 % of forecast 2003 E-commerce B2C sales were estimated to be taxable. States were assumed to collect about 20.9 % of the due revenues through either the sales or use tax, based on the assumptions that all liabilities on automobile sales are collected and 10 % of liabilities on other categories are collected.

Many categories of B2B E-commerce sales were assumed to be exempt, but the largest categories of expected sales were computing and electronics and motor vehicles. The vast majority of both were assessed as taxable in the study. Examples of exemptions in these categories were for custom software and computers used for research in some states and for computers used directly in the manufacturing process. Paper and office products and pharmaceutical and medical purchases were examples of other categories where many purchases are taxable. In total, 52.5 % of expected B2B sales were assumed to be taxable, based on assumptions about what percentage of each of Forrester's categories were taxable.

In some states, certain sales of tangible personal property are taxable but sales of a digital counterpart were not. For example, all states taxed pre-packaged software, but 16 states did not tax software if it were downloaded. No explicit adjustment was made to account for changes in form that alter taxability of transactions, except as assumptions were made about the extent to which certain types of sales were assumed to be taxable. The Fox study concluded that at most one-seventh of E-commerce sales appeared to reasonably fit into the category where they might become nontaxable in some states because they were sold in digitized form. (software, music, books, etc.). The resulting estimates were overstated (assuming that states did not alter the tax base to reflect this trend) to the extent that this shift reduced the tax base, but most

states were assumed to react to such base erosion and redefine the base to include many digitized sales.

Revenue losses from E-commerce equaled taxes due minus use taxes collected. Further, an incremental loss from E-commerce occurred only to the extent that taxes on the transactions would have been collected without E-commerce. These two factors were combined to obtain the final loss estimate. No precise estimates were available on the extent to which use taxes are being paid on B2B transactions. State revenue officials suggested 40 to 50 % compliance is the current average, except for motor vehicles where compliance was much better. The baseline estimates used here assumed 50 % use tax compliance for all items, except for vehicles where the compliance rate was 100 %. This resulted in a weighted average 65.2 % compliance rate. This was assumed to be an upper bound on compliance for E-commerce sales. Also, the baseline assumption used in the Fox analysis was that 50 % of the B2B revenue loss and 35 % of the B2C revenue loss would have occurred even without E-commerce transactions (because of failure to collect sales and use tax in an off-line environment).

Evaluation of the Fox Study

Forrester Projections

Forrester Research (http://www.forrester.com/home) has widely publicized its summary estimates of E-commerce. The underlying data and any analysis are available only to clients, so it is difficult to make full evaluations of the methodology. Figures often cited to support extremely high internet sales were based on $87.5 billion in national consumer E-sales for 2002 suggested by Forrester, who revised them down to $72.1 billion recently. Also, the 2007 figure was revised down from $276.6 to $217.8 billion. Forrester estimated that consumers spent $51.3 billion last year. They anticipated a 40% gain this year, whereas the Census figures show a 22% gain. It does not take many years of growth for differences of this magnitude to result in wildly different projections.

The Fox study was completed in April 2000. At the point in time when the study was conducted, accurate data on E-commerce was not readily available. In addition, projections for future growth of E-commerce sales were high. For example, Forrester projected a rapid compound growth rate of 83.7 % annually through 2003. Based upon more recent survey information from the Census Department, the actual growth in E-commerce sales is substantially below the Forrester projections. While manufacturing has the largest E-commerce component of all industries (18.4 % in 2000), the growth from 1999 to 2000 was only 6.5 %. In fact, the most recent information on E-commerce relating to the Retail Trade came out on August 22, 2002 and on-line sales actually declined in the first two quarters of 2002. Census estimates for E-commerce indicate that the Forrester projections from several years ago were highly exaggerated compared to

what has actually occurred. As a result, any projections based on those exaggerated projections would likewise be much higher than the actual E-commerce data would suggest. Internet sales have demonstrated slower growth more recently.

Identification of the sales taxable components of these sales The categories that Forrester used were relatively broad, so the Fox study made assumptions about the %age of sales for each sales category that would be taxable on average across the U.S. Given that this was done for all 50 states and Washington DC, the assumptions about the %age of sales that were taxable were too general for both the broad categories and for any given state. As a result, even if the Fox methodology were accurate in general, it is unlikely to be accurate for a specific state because of significant differences in the sales tax treatment across states for many intermediate business transactions (B2B sales).

Assumptions about what share of taxable sales could be collected in the absence of E-commerce The baseline estimates used with the Fox study assumed 50 % use tax compliance for all items, except for vehicles where the compliance rate was 100 %. This resulted in a weighted average 65.2 % compliance rate. Also, the baseline assumption used in the Fox analysis was that 50 % of the B2B revenue loss and 35 % of the B2C revenue loss would have occurred even without E-commerce transactions (because of failure to collect sales and use tax in an off-line environment). The Fox assumption is generally consistent with audit results for Minnesota. The baseline assumption that 50 % of the B2B revenue loss and 35 % of the B2C revenue loss would have occurred even without E-commerce transactions, however, is speculative at best.