MINNESOTA Department of Revenue

January 11, 2001

The Honorable Lawrence J. Pogemiller, Chair
Senate Tax Committee

The Honorable Ron Abrams, Chair
House Tax Committee

Commissioner Pamela A. Wheelock
Department of Finance

Dear Senator Pogemiller, Representative Abrams, and Commissioner Wheelock:

As required by Laws 2000, Chapter 490, Article 4, Section 34, alternatives to the required attachment of the federal income tax return to the Minnesota return were studied. The study was conducted in consultation with the Coordinating Committee established in Minnesota Statutes, Section 270.0681.

Enclosed is the report containing the findings of the Coordinating Committee. I concur with the findings of the Committee that none of the alternatives would satisfactorily replace or improve the accessibility and timeliness of the information now provided by the required attachment of the federal return. Therefore, no law change regarding the attachment of the federal return is recommended at this time.

Minnesota Statutes, Section 3.197, specifies that a report to the Legislature must include the cost of its preparation. The cost of preparing this report was approximately $25,000.

Sincerely,

Matthew G. Smith
Commissioner of Revenue

cc: Patrick E. Flahaven, Secretary of the Senate
Edward A. Burdick, Chief Clerk of the House
Marilyn Cathcart, Director, Legislative Reference Library

January 11, 2001

The Honorable Lawrence J. Pogemiller, Chair
Senate Tax Committee

The Honorable Ron Abrams, Chair
House Tax Committee

Commissioner Pamela A. Wheelock
Department of Finance

Dear Senator Pogemiller, Representative Abrams, and Commissioner Wheelock:

As required by Laws 2000, Chapter 490, Article 4, Section 34, alternatives to the required attachment of the federal income tax return to the Minnesota return were studied. The study was conducted in consultation with the Coordinating Committee established in Minnesota Statutes, Section 270.0681.

Enclosed is the report containing the findings of the Coordinating Committee. I concur with the findings of the Committee that none of the alternatives would satisfactorily replace or improve the accessibility and timeliness of the information now provided by the required attachment of the federal return. Therefore, no law change regarding the attachment of the federal return is recommended at this time.

Minnesota Statutes, Section 3.197, specifies that a report to the Legislature must include the cost of its preparation. The cost of preparing this report was approximately $25,000.

Sincerely,

Matthew G. Smith
Commissioner of Revenue

cc: Patrick E. Flahaven, Secretary of the Senate
Edward A. Burdick, Chief Clerk of the House
Marilyn Cathcart, Director, Legislative Reference Library

January 11, 2001

The Honorable Lawrence J. Pogemiller, Chair
Senate Tax Committee

The Honorable Ron Abrams, Chair
House Tax Committee

Commissioner Pamela A. Wheelock
Department of Finance

Dear Senator Pogemiller, Representative Abrams, and Commissioner Wheelock:

As required by Laws 2000, Chapter 490, Article 4, Section 34, alternatives to the required attachment of the federal income tax return to the Minnesota return were studied. The study was conducted in consultation with the Coordinating Committee established in Minnesota Statutes, Section 270.0681.

Enclosed is the report containing the findings of the Coordinating Committee. I concur with the findings of the Committee that none of the alternatives would satisfactorily replace or improve the accessibility and timeliness of the information now provided by the required attachment of the federal return. Therefore, no law change regarding the attachment of the federal return is recommended at this time.

Minnesota Statutes, Section 3.197, specifies that a report to the Legislature must include the cost of its preparation. The cost of preparing this report was approximately $25,000.

Sincerely,

Matthew G. Smith
Commissioner of Revenue

cc: Patrick E. Flahaven, Secretary of the Senate
Edward A. Burdick, Chief Clerk of the House
Marilyn Cathcart, Director, Legislative Reference Library

Study of Alternatives to the Federal Return Attachment for the Individual Income Tax Sample

Findings by the

Coordinating Committee

December 14, 2000

Table of Contents

Introduction

This study was mandated by a law enacted in the 2000 Session of the Minnesota Legislature (Laws 2000, Chapter 490, Article 4, Section 34, reprinted in the Appendix C). The study was the result of a comprehensive effort by the Department of Revenue to reengineer its income tax processing systems and methods. As part of this effort, the Department contemplated not requiring taxpayers to attach a copy of the federal return in an effort to provide efficiency in processing income tax returns and eliminating any burden on taxpayers. However, the federal return provides key information used in developing the state budget forecast and in analyzing the current tax structure and proposed legislation.

The Commissioner of Revenue is required to study and evaluate alternatives to requiring taxpayers to attach a copy of their federal return when filing their Minnesota income tax return. The purpose is to determine whether income tax sample data now obtained from the attached federal return could be obtained by another method and still ensure the continued reliability of the sample data and continued access by the Department of Finance and legislative staff. The study is also to include the evaluation of ways in which the quality of the income tax sample may be improved.

As required by the law, the study was prepared in consultation with the Coordinating Committee established in Minnesota Statutes, Section 270.0681 (reprinted in the Appendix C). The Committee consists of the director of the Research Division of the Department of Revenue who serves as chair of the Committee, the state economist, the chair of the Committee on Taxes of the House of Representatives or the chair's designee, and the chair of the Committee on Taxes and Tax Law of the Senate or the chair's designee.

Executive Summary

Meetings were held at the Department of Revenue on the following dates:

June 7, 2000, 9:00 – 11:00
July 12, 2000, 9:00 – 11:00
August 16, 2000, 9:00 – 12:00
September 19, 2000, 9:00 – 12:00
October 10, 2000, 9:00 – 12:00
October 31, 2000, 9:00 – 12:00
December 14, 2000, 1:30 – 3:00

In addition to the Coordinating Committee, interested parties from the Departments of Revenue and Finance, members of the Legislature, and legislative staff were invited to the meetings.

Frank Martin, Ph.D., Statistical Consultant and Associate Professor in the School of Statistics at the University of Minnesota, was hired as a consultant to assist in the evaluation of the alternatives to the attached federal return. As a recognized expert in the field of statistics, his involvement provided an objective perspective to apply to the charge of the Committee. He also examined the current sampling methodology and made recommendations for improvements to the sample.

Three alternatives to requiring the attachment of the federal return were explored:

Alternative A: Obtain the federal data from IRS tapes, routinely provided to the Department of Revenue.

Alternative B: Request federal returns after the filing of the Minnesota return from taxpayers in the sample.

Alternative C: Instruct pre-selected taxpayers to attach a copy of the federal return, supplemented with a smaller, but significant, post selection process to ensure sample integrity.

See page 13 for a list of all elements considered and addressed by the Committee.

A detailed discussion and evaluation of each alternative are contained in the body of this report.

Recommendation: While the Committee actively pursued and evaluated options for obtaining federal information that is essential for state tax policy and revenue forecasting, none has proven to be adequate to replace or improve the accessibility and timeliness of the information which is currently provided by the legal requirement to attach a federal return for all taxpayers. Therefore, it is the unanimous recommendation of the Coordinating Committee that the law requiring Minnesota taxpayers to attach a copy of their federal return not be changed at this time. The Committee can and will revisit this issue in light of future developments.

The Committee recognizes that the requirement to attach the federal return involves administrative costs for the Department of Revenue and inconvenience to the taxpayer. The Committee finds that the information derived from the federal return is valuable and worth the cost to obtain and maintain it.

In addition to the evaluation of alternatives, the creation of panel data was considered by the Committee. The Committee felt that while the idea had merit, it should be considered and evaluated as a separate issue in the future.

Dr. Martin's recommendations on size and stratification of the sample and computation of confidence intervals for data elements are in the Appendix.

Background

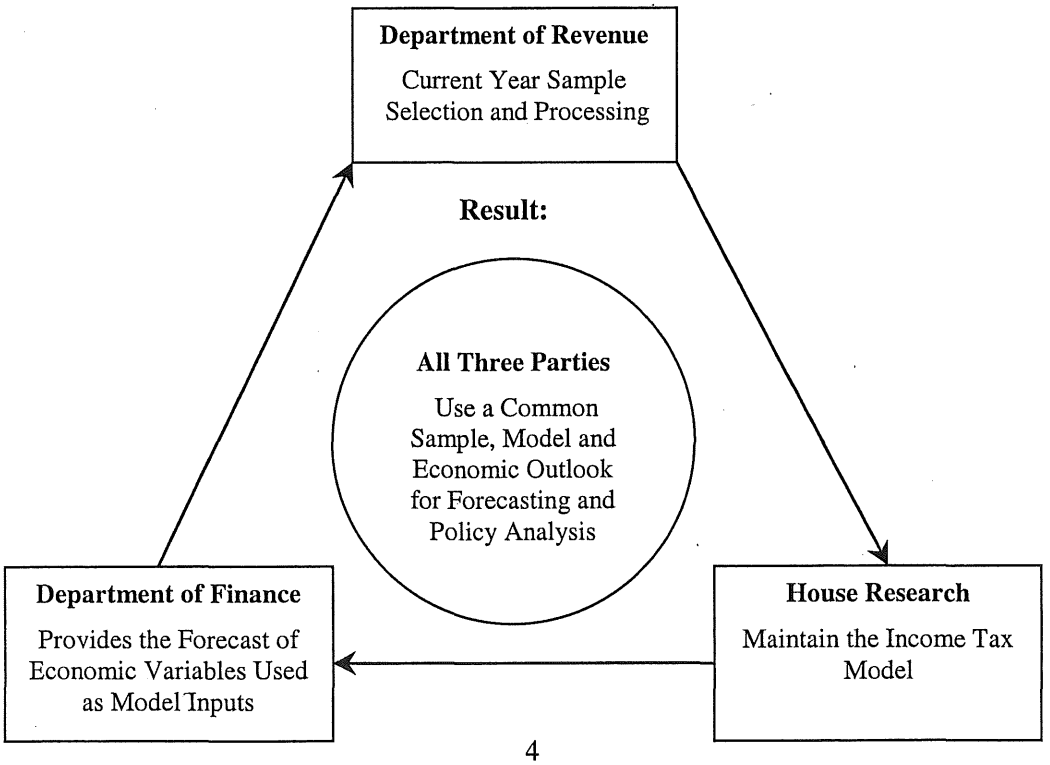
The computation of the Minnesota individual income tax, the state's largest single revenue source, begins with federal taxable income. Rather than requiring the taxpayer to show the computation of federal taxable income on the state return, the law requires a copy of the federal return to be attached when the state return is filed, and federal taxable income is line 1 on the state return. Therefore, many items of information are not repeated on the state return, such as sources of income, components of gains and losses, itemized deductions, and personal and dependent exemptions.

The processing file contains only certain data elements from the Minnesota Form M-1. In order to obtain estimates of the aggregate information from the federal return and the Minnesota schedules, a random stratified sample of state returns is drawn each year. Data elements from the federal return, federal schedules, and state schedules are entered. The additional data is used extensively by Department of Revenue and legislative staff to analyze the current tax structure and proposed law changes. It is also used by the Department of Finance to analyze and forecast state tax revenues.

In addition to using the same income tax sample, the Departments of Revenue and Finance and legislative staff use the same income tax model, which is maintained by staff of the House of Representatives Research Department. The projected rates of change for economic variables in the model are determined by the Department of Finance, so that the model results are consistent with the latest forecast of revenues.

This cooperative effort among the parties involved has resulted in efficient use of resources because costs are shared and there is no duplication of efforts. The use of a common source and basis for analysis and forecasting means that consistent results are obtained by all parties. Therefore, policymakers can focus on policy issues rather than on differences in estimates prepared by legislative and agency staff.

Use of Income Tax Sample Data
In Minnesota State Government



When a copy of the federal return is attached to the state return, the federal return is part of the state return. Therefore, data from the attached federal return is state data, subject to disclosure rules under state law. Before the Department of Revenue gives the sample to the Department of Finance and legislative staff, all identifying information is removed, including name, social security number, address, and any other geographic indicator. Also, other information that could possibly lead to the identification of an individual is blurred.

Need for the Study

The Department of Revenue has undertaken the reengineering of its individual income tax system. The reengineering effort was undertaken with the expressed guarantee that the required federal data would continue to be available to all parties at the same or enhanced level of quality, meeting required timelines. As part of that effort, a recommendation was made to explore ways to eliminate the attachment of the federal return. Under the proposed law change, the statute would give to the Commissioner of Revenue the authority to require the federal return instead of the requirement being in the statute. The federal return would still be considered part of the state return, even if it was not required to be submitted at the time of filing.

The attached federal return is used by many units within the Department of Revenue in administering and enforcing the income tax. However, concern regarding the proposed legislation focused on its potential impact on the income tax sample. Therefore, the 2000 Legislature did not change the statutory requirement but instead mandated this study.

Meetings Conducted

The seven meetings that were held followed basically the same pattern. Updates were provided on the Department of Revenue's reengineering effort, followed by discussion of the alternatives by the Committee with input from the statistical consultant.

Evaluation of Alternatives

The issues evaluated for each alternative were:

- timeliness and coverage of the data
- quality of the data
- availability of the data to personnel outside the Department of Revenue
- resources needed to obtain the information
- efficiency gained or lost compared to the current system
- statistical validity
- response from taxpayers/political considerations

For all of the alternatives, it was assumed that the federal return would always be required at the time of filing for three groups of taxpayers: nonresidents and part-year residents; taxpayers filing after the due date (extension filers); and electronically-filed returns.

Alternative A: Obtain the federal data from IRS tapes

Description
Each year the Department of Revenue receives a tape from the Internal Revenue Service containing information for income tax returns with Minnesota addresses which were processed through the 34th week of the year (plus those processed after the 34th week of the prior year). A supplemental tape is received in December for state returns that did not match with the first federal tape.

The federal information would be available in the time frame needed for returns in the sample, except for nonresidents, extension filers, and those returns that did not match the federal tape for any other reason.

Most of the data elements now entered from the federal return are on the federal tapes. There would need to be some adjustment to the income tax model for the missing data elements, which are listed on Table 1.

Table 1
Tax Year 1998 Income Tax Sample Fields
Not Available on IMF/IRTF

Form	Line Number	Description
Schedule A *Itemized deductions*	7 8 20 21	Personal property taxes Other taxes unreimbursed employee expenses tax preparation fees
Schedule C *Profit or loss from business*	13	depreciation
Schedule D *Capital gains and losses*	6 13 14	short term loss carryover capital gain distributions long term loss carryover
Schedule E *Supplemental* *Income and loss*	28ah 28ak 28bg 28bi 28bj 39	passive income from Sch. K-1 nonpassive income from Sch. K-1 passive loss allowed nonpassive loss from Sch. K-1 section 179 expense deduction net farm rental income or loss
Schedule F *Profit or loss* *from farming*	16 31 35	Depreciation Taxes total expenses

Evaluation

The coverage and quality of the data on the federal tapes were examined. Comparisons were based on tax year 1997, the latest year for which a complete sample was available at the time of the study.

Table 2 shows the results of matching the state income processing file with the first federal tape available for that year and also with the tape that would be available the following year. The federal records on the first tape were matched to 87.3% of the records on the state processing file. An additional 2.3% were picked up from the tape from the following year. Of the 89,000 returns with Minnesota tax that did not match either federal tape, 55,000 were returns from nonresidents/part-year residents. For Table 2, the supplemental federal tape which includes some of the nonresidents/part-year residents was not used.

Table 2
Results of Matching 1997 Minnesota Income Tax Returns with Federal Tapes

	Returns		FTI on MN Return		MN Tax Liability*	
	Number	Percent	Amount	Percent	Amount	Percent
			($ Millions)		($ Millions)	
Matched with						
1st Federal Tape (Rec'd. 9/98)	2,095,296	87.3%	$57,109	71.9%	$4,190	87.8%
2nd Federal Tape (Rec'd. 9/99)	55,777	2.3%	3,933	4.9%	316	6.6%
Total Matched Records	2,151,073	89.6%	$61,042	76.8%	$4,506	94.4%
Unmatched Records	249,542	10.4%	$18,425	23.2%	$265	5.6%
Total Records	2,400,615	100.0%	$79,467	100.0%	$4,771	100.0%
Unmatched Records Only						
With MN Tax > Zero						
Residents (State = MN)	33,953	13.6%	$2,621	14.2%	$154	58.0%
Nonresidents	55,180	22.1%	15,804	85.8%	111	42.0%
Total with MN Tax	89,133	35.7%	$18,425	100.0%	$265	100.0%
With MN Tax = Zero**						
Residents	143,999	57.7%				
Nonresidents	16,410	6.6%				
Total No Tax Returns	160,409	64.3%				
Total Unmatched Records	249,542	100.0%				

* MN tax liability is the tax after nonrefundable credits but before refundable credits. It is equal to line 7 on the 1997 Form M-1.

** Includes returns that were filed only to claim the property tax rebate.

Because most of the unmatched records with tax are nonresidents, Table 3 compares for several variables the totals based on the sample for residents with the totals from the federal tapes for the records matched to the first federal tape. The missing records accounted for only 1% of the returns but 9.4% of the tax. It is assumed that most of the missing records were extension filers. The percentage missing was much higher for some variables, typically the types of income reported by extension filers.

Table 3
Comparison of 1997 Income Tax Data:
Data from Sample for Residents and Federal Tape Data for Matched Records
(Amounts in $ Millions, net of losses)

	Universe From Sample - Residents Only	Processing File Records Matched To 1st Federal Tape	Missing from Matched Records	
			Amount	Percent
Number of Returns	2,116,543	2,095,296	21,247	1.0%
FAGI	$90,987	$85,419	$5,568	6.1%
Sch. A - Item. Ded.	$13,056	$12,359	$697	5.3%
Interest	$2,751	$2,463	$288	10.5%
Dividends	$1,814	$1,575	$239	13.2%
Sch. C - Bus. Income	$3,069	$2,603	$466	15.2%
Sch. D - Cap. Gains	$5,154	$3,949	$1,205	23.4%
Schedule E - Rents, etc.	$5,071	$3,857	$1,214	23.9%
FTI on MN Returns	$60,936	$57,109	$3,827	6.3%
MN Tax Liability	$4,625	$4,190	$435	9.4%

For Table 4, the sample was matched to the first federal tape and then to the supplemental tape received in December. When a federal record matched a sample record, the data elements from the federal tape were substituted for the original federal data in the sample. If there was no match, the original data was not changed. With the substituted data for the matched records, the sample was converted to the universe. Table 3 contains the comparisons for several variables. For all except one item, the variation using the federal tape data compared to the original sample data was less than 0.5%. Of the 24,699 returns in the sample, 3,799 did not match to either federal tape.

Table 4
Comparison of 1997 Sample Results: Original Data and Using Federal Tape Information for Matched Sample Records
($ Millions)

	Original Sample	Using 1st Federal Tape		Using 1st & Supplemental Federal Tapes	
		Amount	% Variation from Original	Amount	% Variation from Original
Wages	$74,975	$74,875	(0.13%)	$74,888	(0.12%)
Taxable Interest	$4,577	$4,570	(0.15%)	$4,568	(0.20%)
Taxable Dividends	$3,026	$3,016	(0.33%)	$2,989	(1.22%)
Net Capital Gains	$12,172	$12,161	(0.09%)	$12,167	(0.04%)
Schedule E Income	$9,027	$9,010	(0.19%)	$9,013	(0.16%)
Taxable Pensions	$4,146	$4,144	(0.05%)	$4,144	(0.05%)
Taxable Social Security	$1,071	$1,072	0.09%	$1,071	-
Federal Adjusted Gross Income	$112,596	$112,463	(0.12%)	$112,453	(0.13%)
Federal Taxable Income	$80,495	$80,518	0.03%	$80,510	0.02%
Total Itemized Deductions	$16,198	$16,218	0.12%	$16,217	0.12%
Total Mortgage Interest	$5,080	$5,083	0.06%	$5,082	0.04%

Matched to 1st federal tape (Sept. 98):	19,921	80.6%
Matched to supplemental federal tape (Dec. 98):	979	4.0%
Unmatched records:	3,799	15.4%
Total 1997 Sample	24,699	100.0%

It appeared, at least in aggregate, that the quality of data on the IRS tapes did not vary greatly from the sample data. Statistics were not produced on variations in individual records. Therefore, it is not known if the aggregate information masked offsetting errors that may have occurred in individual records. The continued quality in subsequent years would depend upon the IRS. It is assumed that most of the unmatched records would be accounted for by nonresidents, part-year residents, and extension filers who would still be required to attach their federal return when filing their Minnesota return.

Some efficiency would be gained in using the IRS tape data because the federal information would not have to be entered for a large number of taxpayers in the sample. The resources currently applied to complete the sample would be shifted to analysis and quality control. The level of compliance for attachment of the federal return by nonresidents and extension filers is not expected to be a significant factor.

The use of the IRS tapes would not affect the statistical validity of the sample.

Although data from a federal return attached to a Minnesota return is considered state data, federal return data on the IRS tapes is considered federal data, subject to disclosure rules under federal law. Currently the Minnesota Department of Revenue uses the federal tape data for administrative and enforcement purposes. The Department also uses the tape data for analytical purposes, releasing only summary information.

In order to substitute federal tape data for data from attached federal returns for use in the sample, the Department of Finance and legislative staff would need to be allowed to have access to the return information under federal rules.

The Committee has concluded that the IRS will not approve the sharing of federal data with persons outside the Department of Revenue that allows for continuation of producing the current income tax sample at a comparable level of quality and reliability.

Conclusion

The issues of the timeliness and coverage of the data on the federal tapes could be dealt with sufficiently in order to use the federal tape data and not degrade the sample. The quality of the data appears to be good, but the maintenance of that quality would be dependent upon the IRS. This alternative would be favorable with regard to efficiency and resources needed to do the sample and would not be expected to cause problems for the statistical validity of the sample or response from taxpayers. Timing also did not appear to be an issue with this alternative. However, the fact that the sample data would not be available to the Department of Finance and legislative staff rules out this alternative.

Alternative B: Request Federal Returns After the Fact from Taxpayers in the Sample

Description
After the sample is selected, taxpayers in the sample who filed state returns would be sent a letter requesting a copy of their federal return. Taxpayers who did not respond to the first letter would be sent a second letter. If still no response, a telephone call and other follow-up work would be needed. The rate of response is not known. A poor response rate may necessitate the selection of supplemental returns and lead to over-sampling in subsequent years.

Evaluation
Under this alternative, the federal data would still be part of the state return and would be available to the Department of Finance and legislative staff. The quality of the data that is captured would likely be similar to the current system.

Additional resources would be needed to contact taxpayers. The initial response rate and the extent to which follow-up work is pursued would affect the level of resources needed. Over-sampling might be needed in anticipation of a poor response rate, resulting in reduced efficiency.

A poor response rate could also affect that statistical validity of the sample if the characteristics of nonrespondents were different from those of respondents.

It is expected that Alternative B could cause significant problems for the timeliness of the sample. Under the current system, the preliminary and final samples are completed at the time they are needed. Under Alternative B, that schedule could probably not be maintained.

Even if a way could be found to deal with the problems mentioned, complaints by taxpayers to their legislators about being "singled out" could put the entire project in jeopardy. Because some income ranges are sampled at 100%, there would be taxpayers who felt they were singled out year after year.

Conclusion
Sample data would be available to personnel outside the Department of Revenue, and the quality of the data should not change. However, Alternative B is ruled out as a viable option for several reasons. It would require significant additional resources. There is no certainty that the sample would be completed when it is needed. The potential negative response from taxpayers and legislators could be such that the sample could not be done in this way in subsequent years, putting the entire process at risk.

Alternative C: Instruct Pre-Selected Taxpayers to Attach a Copy of the Federal Return

Description
The initial sample would be based on the income tax filings for the previous year. Taxpayers in the sample would be selected based on their income and status that year. A letter would be sent to the pre-selected taxpayers at the beginning of the year telling them that they are required to attach a copy of their federal return when filing their state return.

Due to changes in taxpayer income and to the level of response, the pre-selected sample may need to be larger than the target sample. After the state returns are filed, the available sample returns would need to be evaluated against the processing file for the current year to make sure that the sampling rates are satisfied. The sample would also need to be supplemented for first-time filers. The additional taxpayers would be contacted for a copy of their federal return. Follow-up work would be needed for the pre-selected and the supplemental returns.

Evaluation
The federal return data would be available to Department of Finance and legislative staff. The quality of the data could be similar to that under the current system.

The pre-selection would affect the validity of the sample. The fact that they are pre-selected could cause some taxpayers to change how they complete their return. To the extent that this occurred, the sample would not be representative of the population

Pre-selection could also introduce a statistical bias because some types of income, such as capital gains, may be distributed differently by income class in the previous year compared to the current year.

Alternative C has many of the same drawbacks as Alternative B. Additional resources would be needed to obtain the information. Efficiency would suffer because the sample would, in effect, be selected twice – once based on the prior year's filing, then it would be compared to the current year's filing. Obtaining the supplemental returns after the fact could significantly delay the completion of the sample.

It is expected that, under Alternative C, taxpayers would complain to their legislators about being singled out similar to under Alternative B, especially taxpayers who are selected every year. The negative response could put the entire project at risk.

Conclusion
Sample data would continue to be available to Department of Finance and legislative staff, and the quality of the sample data should not change. However, Alternative C is not a viable option for several reasons. Pre-selection would call into question the statistical validity of the sample. Additional resources would be required, and the several additional steps are likely to delay the completion of the sample. It would be the least efficient of the alternatives. The response from taxpayers and legislators could make it politically impossible to repeat the process in subsequent years.

CONSULTANT EVALUATIONS 9/5/00

Current Law

Common Time Lines – Current Law and Alternatives

- Preliminary tax year 1999 sample results delivered in Jan. 2001
- Final tax year 1999 sample results in August/September 2001

Current Law - Review Current Sample Scheme

- Confidence interval total and by data element
- Reduce sampling rates to bring size of sample down to approximately 20,000 returns
- Determine sampling rate by income strata
- Sampling rates for residents vs. nonresidents
- Filing status stratification

Alternatives to Required Attachment of the Federal Return:

C1. Obtain Federal Data from IRS Tapes

- Use federal data from IRS tapes for returns of MN residents processed through the 34th week.
- Require nonresidents and taxpayers filing after April 15th to attach a copy of the federal return.

C2. Request Federal Returns for Sample

- Request a copy of the federal return after the fact from taxpayers selected in the sample drawn from the processing file,
- Requires contacting 20,000 to 25,000 taxpayers.
- If no response, how many additional requests should be made?
- What types of problems or bias would be introduced?

C3. Instruct Pre-selected Taxpayers to Attach Federal Return

- Pre-select taxpayers in the sample based on prior-year information.
- Instruct pre-selected taxpayers to attach a copy of their federal return when they file their state return.
- To supplement the pre-selected sample, would need to request federal returns after the fact from ___? number of taxpayers.
- What types of problems or bias would be introduced?

Common Questions

- What stratifier
- Rates/size
- Confidence interval(s)
- How do we treat bad observations
- Other recommendations
- SOI - blurring
- Use of sample for compliance measures

Panel

Incorporation of panel concepts into current law (B) and alternatives (C1, C2, C3)

Sample Related Questions

- Audit rates for using sample for compliance measurement.
- Related sample rates for incidence database.

Recommendations to the Coordinating Committee on Federal Return Attachment

Frank B. Martin, Ph.D.
Statistical Consultant

November 14, 2000

The Sampling Alternatives

The current sample is in large part drawn by physically locating and pulling the state income tax returns with the federal return attachment from the huge file of Minnesota returns. Three alternatives to the process of collecting the file of federal attachments to all Minnesota returns are reviewed below.

Alternative A

Using the IRS tapes as the sampling frame is very desirable if all the data elements required are accurately recorded on the tapes. The frame is complete and would greatly facilitate the current sampling procedure realizing substantial cost savings. The electronic frame would make the drawing of special need samples a viable and efficient idea. It is my understanding that the state does not exercise control over the availability of the federal tapes and cannot currently rely on them for planning and drawing the sample. Alternative A must yield to Alternatives B and C.

Alternative B

Permits drawing the sample from exactly the population of interest and will not introduce sampling biases. However, nonresponse bias may be expected to be a serious problem which large expenditures and great effort will not completely overcome. Considerable taxpayer inducements to supply the federal return will probably produce tolerable results. In some strata taxpayers will have a very much higher probability (a virtual certainty in some cases) of being drawn from year to year and this may lead to unequal treatment complaints. These response bias and unequal treatment issues apply equally to Alternative C.

Alternative C

Suffers from the fact that the prior year's returns do not define the sampling frame for the current year's sample. You would not permit the inclusion of first time filers and you would permit the drawing of last time returns which will not "return" for the current year's sample. These problems are not fatal from a statistical point of view but they do make Alternative C less desirable than B.

Conclusion: Alternative B is most reliable and statistically most desirable but not without practical problems.

Definitions

"MNFDTXBL" is the federal taxable income found on line 1 of the Minnesota tax return.
"FAGI" is federal adjusted gross income found on line 33 of the federal tax return.

FAGI

The sample is currently stratified using MNFDTXBL bins as the strata. A stratifier is preferable if it does a good job of splitting the population of responses up into discrete lumps so that the population variability can be explained by the differences between stratum bins. A stratifier variable is superior if we can get uniformity (small response standard deviations) within the strata and separation between the strata. The better a stratifier is at this task the more closely it can be said to be related to the response. The direct effect of reduced standard deviation within strata is reduced uncertainty (margin of error width) in population estimates inferred from the sample.

Examination of the relationships between FAGI and MNFDTXBL stratifiers and certain important responses such as wages, schedule C income, and capital gains shows FAGI to be the hands down winner over MNFDTXBL. Using FAGI as a stratifier instead of MNFDTXBL will produce margins of error approximately 60% the size of what MNFDTXBL is currently producing.

The simple fact is that FAGI is more closely related to important response variables. When the opportunity presents itself the committee should request the inclusion of FAGI or the substitution of FAGI for MNFDTXBL in the set of variables available at the sample designing stage.

Stratification

Both filer status and residential status have profound effects on the configuration of variables in a return. The 3 x 2 combinations of filing status by residential status should be the primary or first cut stratifiers. The 3 filing status bins recommended are 1) single, 2) joint and widower, and 3) separate and head of household. These bins are further split into resident and nonresident filers. In each of these six bins a further stratification can be customized to the bins characteristics using the unfortunate choice of MNFDTXBL as the stratifier.

An approximately optimal allocation of sample sizes to approximately 50 stratum bins will be constructed with some oversampling in the nonresident separate and head of household bins which constitutes only three tenths of one percent of the population of returns. A stratum of particular interest to legislative proposals is joint returns with low income. Examination of the data from last year's sample suggests that a good definition of this stratum will be MNFDTXBL between $-25,000 and $30,000. This stratum will contain approximately 20% of all returns and will have an assigned sample size of over 3,000 returns.

An additional factor worthy of consideration as a stratifier is electronic vs. paper return. The cost of sampling electronic returns is apparently vastly lower which permits more intensive sampling of electronic returns. There may also be material differences between electronic and paper returns. There is currently a large flux in the size of the electronic population with projected targets waiting to be reached. It is suggested that systematic investigation of the use of this stratifier be delayed for a few years.

Data Issues

A total sample size of 20,000 returns plus "specials" appears to be adequate for the various analytical investigations the sample is employed in.

The proposed sample is large enough so that returns found to have entries corrupted in processing can simply be discarded.

An algorithm for calculating the margin of error on any summary estimate produced from the sample can be constructed using the approximately 50 stratum counts, the stratum sample sizes and the data collected in each stratum.

Panel Construction

It has been suggested that a panel of returns be observed over a period of years to determine how persistent certain income features are, such as capital gains. A convenient way to do this would be to draw a cross section of the tax sample consisting of about 1,000 returns which have features being studied. You could possibly pull the prior year return and begin making comparisons. In each subsequent year an additional 1,000 can be pulled until a set of approximately 5,000 was in the panel. Thereafter, additional returns can be added to compensate for attrition and the 5,000 panel size can be maintained.

Appendices Attached

Appendix A

Appendix A contains the sample sizes for the 49 strata totaling 17,000 allocated to the samples in a nearly optimal manner for the purpose of attaining small margins of error on estimates of universe totals.

Appendix B

Appendix B contains the algorithms for computing the 95% confidence limits for universe totals estimated from the tax sample.

Appendix A

Department of Revenue Income Tax Return Sample

Sample Size for Each Stratum

Total Sample Size = 17,000 Plus Total Number of Returns from 100% Strata
Based on Use of MNFDTXBL

Resident Single Filers: Number Sampled = 5,400

Bin Boundaries	Sample Size	Stratum Code
Special	100%	100
$ -500,000 < $ -25,000	52	101
-25,000 < 0	205	102
0 < 10,000	1,680	103
10,000 < 30,000	1,698	104
30,000 < 50,000	759	105
50,000 < 100,000	540	106
100,000 < 250,000	272	107
250,000 < 500,000	115	108
500,000 < $1,000,000	79	109
$1,000,000 Plus	100%	110

Resident Married Filing Jointly and Widow(er)s:
Number Sampled = 9,600

Bin Boundaries	Sample Size	Stratum Code
Special	100%	200
$ -500,000 < $ -25,000	52	201
-25,000 < 30,000	3,111	202
30,000 < 50,000	1,697	203
50,000 < 100,000	2,337	204
100,000 < 250,000	1,412	205
250,000 < 500,000	597	206
500,000 < 1,000,000	394	207
$1,000,000 Plus	100%	208

Appendix A (Cont.)

Resident Married Filing Separately and Head of Household:
Number Sampled = 860

Bin Boundaries	Sample Size	Stratum Code
Special	100%	300
$ -500,000 < $ -25,000	5	301
-25,000 < 0	86	302
0 < 5,000	201	303
5,000 < 10,000	73	304
10,000 < 20,000	133	305
20,000 < 30,000	104	306
30,000 < 50,000	111	307
50,000 < 100,000	73	308
100,000 < 250,000	44	309
250,000 < 500,000	17	310
500,000 < 1,000,000	13	311
$1,000,000 Plus	100%	312

Nonresident Single Filers
Number Sampled = 240

Bin Boundaries	Sample Size	Stratum Code
Special	100%	400
$ -500,000 < $ -25,000	5	401
-25,000 < 10,000	26	402
10,000 < 50,000	72	403
50,000 < 100,000	18	404
100,000 < 250,000	17	405
250,000 < 500,000	15	406
500,000 < 1,000,000	18	407
1,000,000 < 10,000,000	69	408
$10,000,000 Plus	100%	409

Appendix A (Cont.)

Nonresident Married Filing Jointly and Widow(er)s
Number Sampled = 720

Bin Boundaries	Sample Size	Stratum Code
Special	100%	500
$ -500,000 < $ -25,000	5	501
-25,000 < 30,000	64	502
30,000 < 100,000	125	503
100,000 < 250,000	80	504
250,000 < 500,000	72	505
500,000 < 1,000,000	82	506
1,000,000 < 10,000,000	292	507
$10,000,000 Plus	100%	508

Nonresident Married Filing Separately and Head of Household
Number Sampled = 180

Bin Boundaries	Sample Size	Stratum Code
Special	100%	600
$ -500,000 < $ -25,000	5	601
-25,000 < 20,000	41	602
20,000 < 50,000	11	603
50,000 < 250,000	22	604
250,000 < 500,000	10	605
500,000 < 1,000,000	16	606
1,000,000 < 10,000,000	75	607
$10,000,000 Plus	100%	608

Appendix B

Procedures for Interval Estimation of Universe Total Using the Tax Return Sample

Notation, Definitions, and Computational Steps

h A 3-digit code for strata bins used in Appendix A: for example $h = 202$ defines resident joint filers and widowers in the range -25,000 < 30,000.

Nh The total number of returns in stratum h in the universe of all taxed returns: for example in 1998 $N(202) = 386{,}594$.

nh The size of the tax sample assigned to strata bin h as taken from Appendix A: for example $n(202) = 3{,}111$.

wh The weight factor for every sampled return in stratum h used to inflate the sample total to the universe total: for example, in 1998 wh for $h = (202)$ would have been 386,594/3,111, $wh = Nh/nh$, $w\,(202) = 124.27$ in 1998.

yhi Is the response measured on sampled return i in stratum bin h.

yhi may be a number read right off the tax return or a calculation using numbers read off the tax return.

yhi can also be an indicator variable taking on the values $yhi = 1$ or $yhi = 0$ to tally whether the return possesses ($yhi = 1$) or does not possess ($yhi = 0$) a certain attribute: for example, $yhi = 1$ if return i in the stratum h sample claims a disability.

The sampling frame or universe consists of all tax returns which are not in the exceptional strata where a total census of returns is conducted. In Appendix A, 12 of the 61 strata are in the exceptional category where sampling is 100%. The sampling universe will be defined as the remaining 49 strata.

$N = \sum_{h}^{49} Nh$ is the size of the sampling universe where the summation is over 49 included strata.

$\bar{y}h = \sum_{i=1}^{nh} yhi/nh$ is the mean response in stratum h. In the case $yhi = 1$ or 0, this is the proportion in the stratum h sample which have the attribute.

Appendix B (Cont.)

Notation, Definitions, and Computational Steps (Cont.)

$\hat{Y} = \sum_{h} Nh\ \bar{y}h$ is the estimate of the sampling universe total for response yhi. In the case $yhi = 1$ or 0, $\hat{Y}$ estimates the total number of returns in the sampling universe which have the attribute.

$\hat{Y} = \sum_{h}\sum_{i} (wh)\,(yhi)$ is an alternative computation of $\hat{Y}$.

$$sh = \left[\frac{\sum_{i=1}^{nh} (yhi - \bar{y}h)^2}{nh - 1} \right]^{1/2}$$ is the standard deviation of the sample of nh values of yhi in stratum h.

$(sh)^2$ is the variance estimate of the universe of responses in stratum h.

$V(\hat{Y}) = \sum_{h} Nh(Nh - nh)\ \frac{(sh)^2}{nh}$ is the estimated variance of the $\hat{Y}$ estimate.

$SE(\hat{Y}) = \sqrt{V(\hat{Y})}$ is the estimated standard error of estimate of the sampling universe's total for response yhi.

$C =$ the calculated total for response if yhi in the 12 strata where census was employed.

$\hat{T} = \hat{Y} + C$ is the estimated total for the universe of all tax returns.

$\hat{TU} = \hat{T} + 2\,SE(\hat{Y})$ is the upper 95% confidence bound on universe total T.

$\hat{TL} = \hat{T} - 2\,SE(\hat{Y})$ is the lower 95% confidence bound on universe total T.

Procedures for Estimation of Totals in a <u>Subdomain</u> of the Universe

A subdomain could be a certain subset of the universe of all returns: for example, all returns which report farm income. There are two techniques available for estimation for a subdomain total. The returns which fit into the subdomain may be easily identified and the stratum counts may be readily available or estimation may have to be done without benefit of strata sizes in the subdomain.

Appendix B (Cont.)

Stratum Counts Known

Mh = Number of returns in stratum h which are known to form the subdomain.

$M = \sum_{h} Mh$ total number of returns in the sampling universe subdomain after removal of the 12 censused strata.

mh = number of subdomain returns in stratum h which are found in the sample $mh \leq nh$.

If $mh = 0$ but $Mh > 0$ then an adjacent stratum must be collapsed with stratum h. This can continue until $mh \neq 0$.

$\bar{y}h = \sum_{i=1}^{mh} \frac{yhi}{mh}$ as before.

sh = the standard deviation of the set of mh sampled responses in stratum h. If it was necessary to collapse 2 strata to get $mh \neq 0$ there is a slight problem with unequal sampling rates in the different collapsed strata but this is negligible.

$(sh)^2$ = subdomain stratum variance estimate.

$\hat{YD} = \sum_{h} (Mh)\, \bar{y}h$ is the estimate of the subdomain total in the sample universe.

$$V(\hat{YD}) = \sum_{h} Mh\,(Mh - mh)\,\frac{(sh)^2}{mh}$$

$$SE(\hat{YD}) = \sqrt{V(\hat{YD})}$$

CD = the calculated total of response yhi for the subdomain in the 12 census strata.

$TD = \hat{YD} + CD$

$TDU = TD + 2\,SE\,(\hat{YD})$

$TDL = TD - 2\,SE\,(\hat{YD})$

Appendix B (Cont.)

Stratum Counts (Mh) Unknown

The condition may arise where there is interest in a subdomain defined by a characteristic which is not immediately recorded in the population of all tax returns. This would make determination of the stratum counts (Mh) for this subdomain either impossible or prohibitive.

The solution to subdomain estimation is a definition of the response variable:

y^*hi = yhi if the sampled return is found to be in the subdomain based on its characteristics.
= 0 if the sampled return is outside the subdomain.

Estimation proceeds on the full sample of nh from each stratum of size Nh and all estimates are calculated using the entire sample with summation over the entire sample on this modified response variable.

Appendix C

Laws 2000, Chapter 490, Article 4,

Sec. 34. [TAX INFORMATION SAMPLE DATA STUDY.]
(a) One of the goals of a reengineered income tax system is to reduce the administrative burden for both taxpayers and tax administrators. In order to reduce the cost of handling paper returns and to explore electronic options for taxpayer filing of tax data, the department of revenue will explore eliminating the requirement of Minnesota Statutes, section 289A.08, subdivision 11, that the federal return be attached in filing a Minnesota individual income tax return. This federal return information is used for the purposes of ensuring the accurate calculation of individuals' Minnesota income tax liabilities and for the purposes of preparing the microdata samples under Minnesota Statutes, section 270.0681.
(b) To ensure the continued reliability of income tax data samples and to evaluate ways in which the quality of samples may be improved, the commissioner shall study and evaluate alternatives to requiring taxpayers to attach a copy of their federal return when filing Minnesota state income tax. The study must be prepared in consultation with the coordinating committee established in Minnesota Statutes, section 270.0681, subdivision 2. The study must:

(1) evaluate the quality of federal electronic data compared to sample data prepared from returns filed with the department;
(2) evaluate alternative sampling methodology, including preselection of sampled returns, panel data, and other sampling methods; and
(3) evaluate and test whether alternative methods can
    (i) provide a data sample that is as accurate and reliable as one prepared from federal returns that are filed with or attached to Minnesota individual income tax returns; and
    (ii) result in a data sample that will continue to be available to staff of both the department of finance and the legislature on the same basis as one prepared from returns required to be attached to or filed with the Minnesota tax returns.

(c) The commissioner of revenue shall report the findings of the study to the house tax committee chair, the senate tax committee chair, and the commissioner of finance.
(d) The commissioner of revenue shall, with the approval of the commissioner of finance, prepare a bill for introduction in the 2001 legislative session that eliminates, for some or all taxpayers, the requirement that a copy of the federal return be filed with the individual income tax return, if the commissioner determines as a result of the study that:

(1) an alternative method would provide a data sample that is as accurate and reliable as one prepared from federal returns required to be filed with the Minnesota return; and
(2) the sample will continue to be available to the staff of both the department of finance and the legislature on the same basis as one prepared from returns required to be filed with Minnesota tax returns.

Minnesota Statutes, Section 270.0681 [TAX INFORMATION SAMPLE DATA],
Subdivision 1. Preparation of samples. The commissioner of revenue shall prepare microdata samples of income tax returns and other information useful for purposes of estimating state revenues, (2) simulating the effect of changes or proposed changes in state and federal tax law on the amount of state revenues, and (3) analyzing the incidence of present or proposed taxes.

Subd. 2. Coordinating committee. A coordinating committee is established to oversee and coordinate preparation of the microdata samples. The committee consists of (1) the director of the research division of the department of revenue who shall serve as chair of the committee, (2) the state economist, (3) the chair of the committee on taxes of the house of representatives or the chair's designee, and (4) the chair of the committee on taxes and tax laws of the senate or the chair's designee. The committee shall consider the analysis needs and use of the microdata samples by the finance and revenue departments and the legislature in designing and preparing the samples, including the type of data to be included, the structure of the samples, size of the samples, and other relevant factors.