Microsimulation Modeling in the Russian Federation

Robert McNab
Sally Wallace

Introduction

Microsimulation models, analytic computer models based on individual (micro-level) data observations, have become increasingly popular tools for policy analysis.[1] The increased usage of these models is evident world-wide, and is not confined to developed countries. While the U.S., Canada, Germany, and France have used these models for a number of years, countries such as Jamaica, Guatemala, China, and the Russian Federation have more recently developed microsimulation models for policy analysis.

The reason for the increased usage of these models is clear. Microsimulation models are the best tool for analyzing the revenue impact of changes in revenue policies, i.e., the tax code. This is true because the models can be used to estimate not only the aggregate revenue change, but they can also produce the change in taxes paid by type of taxpayer (firms by firm size, industry, regional location; individuals by income classification, region, or demographic group); they can produce regional analyses of the level of revenue change; they can produce a forecasted revenue stream under alternative tax regimes; and they can be integrated with macroeconomic models in such a way as to show the macroeconomic implications of tax law changes. As data have become more available and computerization has become more widespread, the ability to develop these models has also grown.

The ability to produce the types of analyses listed above makes the models crucial to the work of policy makers, law makers, and tax administrators alike. Policy and law makers are not only interested in the level of revenue change associated with a particular piece of tax legislation, they are also interested in how such a change impacts the economy, particular types of firms, and people. Tax administrators can use the outputs from microsimulation models as an indication of changes in tax compliance behavior under existing or new tax regimes. Data bases for such models can also be used to help develop audit selection criteria.

To date, the Russian Federation has not made significant use of microsimulation models. In large part, this is due to the fact that the data requirements for the models are quite involved and such data are not readily available in many countries, including the Russian Federation. Since the data are based on individual firm or personal observations, questions of confidentiality may also arise. However, in the last two years, significant gains have been made in the development of a microsimulation model for the city of Moscow.

---

[1] This includes both revenue and expenditure policy. The main focus in this chapter is on the revenue applications of microsimulation models.

The structure of the chapter is as follows. In the next section, we explain the methodology behind microsimulation models, in a policy context. In the second section, we present an in-depth discussion of the Moscow City microsimulation model. The third section summarizes some results for a specific tax policy change and the final section discusses future microsimulation model development.

Methodology: Microsimulation Models

Microsimulation models are inherently tools of revenue estimation. That is they serve as a tool which uses microlevel data to analyze the effects of different types of policies on individuals, firms, program recipients, etc. As explained below, these models are also directly and indirectly tools of revenue forecasting whereby the micro level data are projected or extrapolated into the future and total revenues are calculated for these future years.

Microsimulation models have a long history in policy analysis. The methodology behind microsimulation models is based on the work of Orcutt (1957) and Orcutt, Greenberger, Korbal, and Rivlin (1961). The uses of these types of models have been extended to many different types of policy issues across many countries throughout the last three decades. The specific uses of microsimulation models range from estimating the distributional impact of changes in the taxation of social security benefits (US Social Security Administration and Wixon, et. al., 1987), to the demand for day care services in Denmark (Baekgaard, 1996) to the implications of sales tax reform in Canada (Kapur, Grupta, and McGirr, 1997).

There is no one unique approach to microsimulation modeling, however, we can classify microsimulation models into two general types: *static* and *dynamic*. *Static* models are used most often to simulate the short-term potential impacts of detailed changes to tax and transfer programs. *Dynamic* models, on the other hand, are often used to simulate the long-term impact of changes in tax and transfer programs. The distinct difference between the two approaches is that static models assume that personal and enterprise behavior remains unchanged, that is, behavioral decisions are static and can not change in response to changes in tax and transfer programs. Dynamic micro-simulation models try to capture the behavioral responses to changes in tax and transfer programs, and attempt to capture the changes in consumer and producer demands induced by the changes in the tax and transfer structure.

A typical microsimulation model is comprised of three pieces: (1) a micro-level database (typically information from tax returns for individuals or corporations for the base year and future years), (2) a tax calculator (computer program that calculates the tax paid under alternative tax structures and which may be supplemented with "behavioral changes" associated with the tax changes), and (3) an output program which categorizes taxes paid by income group, tax burdens, "winners and losers," and the overall change in revenue. Figure 1 shows the basic structure of a microsimulation model for corporate income tax analysis.

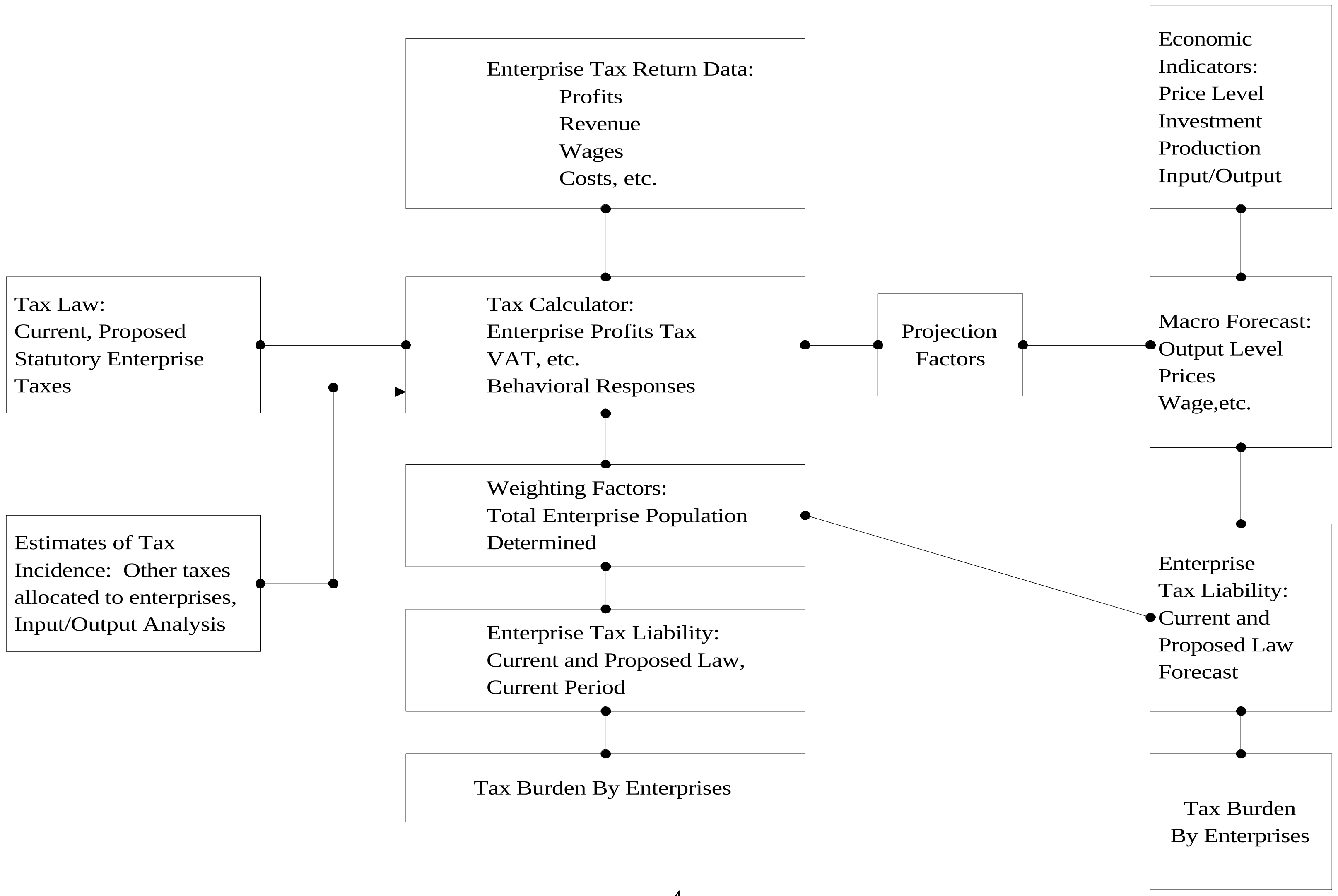
While microsimulation models can be used to analyze a number of policies, here we concentrate on the methods and uses of microsimulation models for analysis of tax policy. The benefits of microsimulation models are: (1) the models provide detailed estimates of revenues by income group, industry group, and type and level of income, (2) the models enable analysis to determine leakages in collections by type of industry, income, and income class, (3) the models provide detailed information on tax bases so changes in the tax law are relatively easy to incorporate (4) the models can provide a consistent link among various taxes--especially income taxes, excises taxes, and sales and property taxes, and (5) the models can incorporate a macroeconomic feedback effect component which allows one to analyze the effect of tax changes on the macro economy and vice versa.

As seen in Figure 1, the core of the model is the micro (firm-level) data which serves as the observation units of the model. These data could just as easily be for individuals. The data are fed into a tax calculator, which contains current and any proposed law tax code. Estimates of behavioral responses could also be incorporated so that the impacts of the tax changes become *dynamic* rather than *static* which is the case when no behavior is included. Weights are applied to the sample observations and finally a variety of output is produced by the model. The right hand side of the figure shows the forecasting feature of a microsimulation model. Many microsimulation models do not contain this forecasting feature, although it is somewhat more common now than in the previous 10 years. The following provides more detail on the components of the model.

Micro-level database. The data for microsimulation models should come from a sample of taxpayers -- individuals or corporations -- and contain all available information from the tax returns or other reporting documents such as annual sales tax reconciliation reports or property tax records.[2] These data should include a sample of residents as well as non-residents whether firms or individuals. The sampling procedures used for database development should be sophisticated enough to ensure an accurate reflection of the true population. Unfortunately, there is no formula to determine the best sampling strategy when we are sampling actual tax return or account information. If the population (whether firms or individuals) is relatively homogeneous, then a basic random sample of population is probably sufficient to ensure an accurate sample. If, however, the income distribution is highly skewed, or certain industries are a much larger component of the population, or the distribution of tax liability is skewed by size of firm, it becomes much more important to institute various stratifications of the data before sampling. The actual type of stratification is largely determined by the characteristics of the population. For example, in the U.S., individual income tax returns with adjusted gross income (AGI) over $200,000 are sampled at a rate 4 times (and more) higher than returns with AGI less than $200,000. Most countries that use microsimulation models for corporate income taxes use a stratified random sampling technique, where the stratifications are made based on criteria such as industrial sector, size of tax liability, total revenue, and assets.

[2] The use of a random sample of individual observations reduces computer storage and processing requirements.

Figure 1: Sample Corporate Income Tax Microsimulation Model Structure



Additional data may be imputed to each observation. For example, it may be useful to "know" the consumption patterns of the taxpayers as well as their individual income tax liability. This would allow analysis of the effects of changes in consumption-based taxes, as well as direct taxes, on individual taxpayers.

Imputing the additional information can be accomplished in either of two ways. One way is to match individual income tax records (observations) with like observations from a micro-level data set that contains information on consumer expenditures. Since one does not know the actual people in both data sets, the matches must be done based on similar characteristics such as income, age of head of household, and family size. Once this match is made, the individual observations contain information on tax liability and also information on the taxpayer's consumption of consumer goods.

A second imputation approach uses regression analysis to determine the relationships among variables on the tax return (income, number of dependents, marital status, for example) and consumption. For main expenditure categories, regressions of the following general form can be estimated using data from the consumer survey:

$$Exp_i = \beta_{0i} + \beta_{1i} INCOME + \beta_{2i} FAMILY\ SIZE + \beta_{3i} MARITAL\ STATUS + \varepsilon_i$$

Once the equations have been estimated, the coefficients from those regressions can be used with the tax data to impute consumption of each type of good. Therefore, this approach also yields a micro-level database with tax and consumption information.

The last component of the micro-data is the projection or aging component. The micro-level data will always be based on some past year, due to data availability. To forecast the revenues under current law (the existing tax structure), or under a new tax proposal, the data need to be aged or extrapolated into the future.

One of the most straightforward ways to age the data separates major income categories, consumption groups, etc. and makes projections of those groups based on past performance, macro-economic indicators produced elsewhere in government or the private sector, and regression analysis. For example, in Russia, the largest component of reported income for the individual income tax is wages and salaries. A time-series regression of wages against GNP, inflation, and employment may yield a stable relationship for wages over time.[3] The coefficients of this regression could then be used with the official macro-economic forecast of GNP, inflation and employment to project the total level of wages and salaries. This would be done for other important categories of income as well. The overall increase in wage and salary income is then attributed across the income groups in the micro-data file.

---

[3] This baseline data forecasting is closely tied to forecasting techniques used to project tax revenues. For more detailed information on these forecasting techniques, see the forecasting chapter in this volume.

Once the aged data are available, the sample of tax filers is then representative of the tax paying population in some future year. The computed tax liabilities using the aged data weighted to the population totals represent the forecast of actual liabilities. These liabilities can be calculated under current law or proposed law depending on the tax calculator.

Using microeconomic tax return data raises the important issue of taxpayer confidentiality. In the case of the joint Moscow City STI-Georgia State University (GSU) microsimulation model development, issue surfaced quickly. With respect to the confidentiality of the data, there were three general approaches that were considered during the development of the Moscow City Model. The first approach was to release all the data, including identifying information, to GSU with the proviso that GSU would assume legal responsibility for the confidentiality of the taxpayer data. The second approach was to release the data without identifying information but with the Taxpayer Identification Number (TIN). The inclusion of the TIN would enable the Moscow STI and GSU to check the calculations of the micro-simulation model without having to re-process the data. The third approach was to release the data without identifying information but with a transformed TIN. The third approach is the most secure with respect to the confidentiality of the data, however, the Moscow STI would have to keep a master list that would allow the matching of the transformed TINs with the actual TINs. In the end, the decision was made by the Moscow STI-GSU project team to work with taxpayer data that contained transformed TINs. This issue can be overcome in most countries, but there are different levels of tolerance in each country.

Tax Calculator. This part of the model is a straightforward computer program that calculates the actual tax liability for each individual observation in the data set. The computer program is basically a series of statements reflecting the calculations that taxpayers actually make when computing their tax liability. By changing parts of the tax calculator to reflect changes in the law, the program can "simulate" new tax liabilities for any number of proposals.

The tax calculator can also assign taxes based on various tax incidence assumptions. For example, if the burden of the company income tax was assumed to fall 50% on labor and 50% on capital, the program could allocate tax changes in accordance with these incidence assumptions (which could also be easily changed).

Within the context of the tax calculator, more sophisticated behavioral effects can be incorporated. For example, if there is evidence based on past experience and empirical study that increases in the corporate marginal tax rate resulted in lower real wages (due to the corporate tax being shifted), a behavioral response (elasticity) could be “turned on” to reduce the wage payment of the corporations as a function of their change in marginal tax rate. Other behavioral responses could be changes in the level of investment, output, and other deductions. The problem with incorporating these behavioral responses is that even in countries with relatively stable tax systems and lots of historical data, researchers have found it hard to agree on the magnitude and direction of these types of responses. Symons and Warren (1996) show the importance of behavioral responses in their study of commodity tax reforms in Australia. They discuss the difficulties of adding such behavior and stress the importance of keeping the model tractable. Klevmarken and Olovsson

(1996) demonstrate the potential magnitude of behavioral impacts in their study of income tax changes in Sweden. They conclude that while worth the effort, their revenue impacts of including behavioral changes are relatively small; although the impacts on the distribution of income may be quite significant. In a country such as Russia, where the historical data needed to estimate such responses is virtually non-existent, it would be very difficult (and possibly misleading) to incorporate such behavioral responses. As time goes on and more historical data are developed, these responses could be incorporated. Also, it is possible to use the experiences and information from behavioral responses from other countries, although one must be careful to determine the applicability of such "outside" information.

Table 1 shows an example of how a tax calculator would quantify a change in the tax rate and the standard deduction for the Personal Income Tax. In this example, a proposed change in the tax law would decrease the standard personal deduction from 3,500 to 1,500 and reduce the tax rate from 30 percent to 20 percent to compute the new tax liability. The calculator simply uses data available on the tax declaration to make the current law and proposed law calculation. The fundamental step in a microsimulation model is to perform the same computation to a representative sample of the taxpayers to determine the overall impact of the proposed change in the tax structure. Likewise, the same type of methodology can be applied to the Enterprise Profits Tax, the Value Added Tax, or other taxes of interest as long as the information is available to replicate the tax structure.

Microsimulation models are invariably more complex than the example in Table 1. For example, the Russian tax code is much more intricate than implied in Table 1, with multiple tax rates at different income levels, itemized deductions, and excluded sources of income. To simulate one tax actually involves many lines of computer code, to make the model "user friendly" involves a significant increase in the amount of code. User friendly models create a tool which allows analysts to easily change major components of the tax code such as rates, exemptions and deductions.

TABLE 1

SIMULATION OF PERSONAL INCOME TAX RATE REDUCTION: SINGLE PERSON

Variables	Current Law	Proposed Law
Income		
Gross Revenues	60,000	
Interest	500	
Dividends	2,000	
Total Income	62,500	
Deductions		
Standard Deduction	3,500	1,500
Per Person Deduction	2,000	2,000
Total Deductions	5,500	3,500
Taxable Income	57,000	59,000
Tax Rate	30 percent	20 percent
Tax Liability	17,100	11,800
Average Tax Rate (Taxes Divided by Total Income)	27.36 percent	18.88 percent

The design and programming of the microsimulation models are dependent upon the needs of the user and the resources of the programmer. The model should be readily accessible to another programmer and in a common programming language. The model should not require major changes to the code in the future if the structure of the database changes or if the tax laws are reformed. Finally, the microsimulation model should have the flexibility to allow the analyst to examine current topics of interest but also future topics of interest.

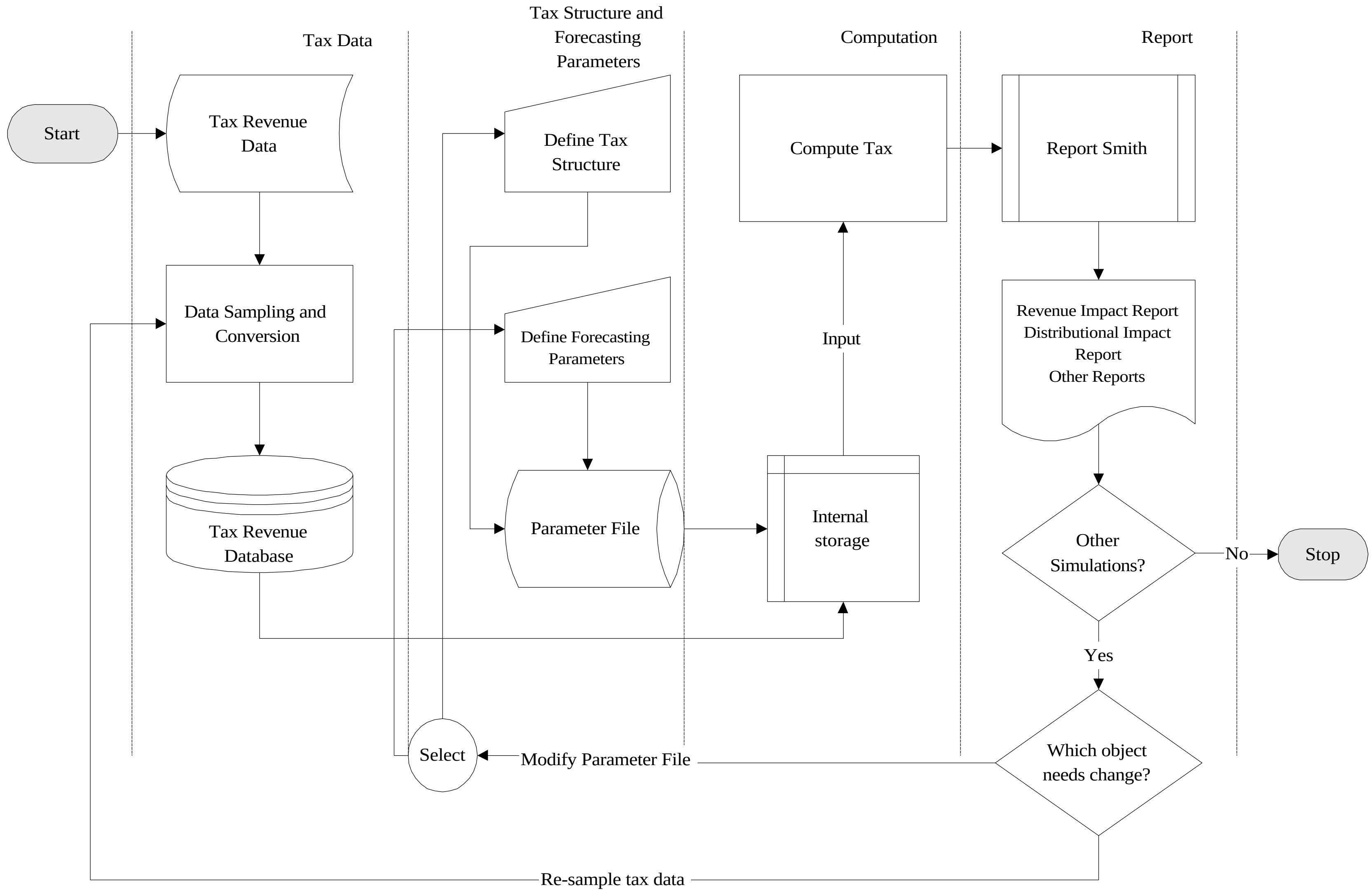
The model developed for Moscow City attempted to address each of these objectives by adapting a user-friendly interface where the underlying programming code is a well known language. The language is *Delphi*, and Object Oriented Programming language, which allows the programmer to build GUI-based interfaces for ease of use while retaining the powerful properties of the Pascal language. Since *Delphi* uses Object Oriented Pascal, programmers from the Moscow STI can modify the core programming code without the need of outside consultants. Since the model is not "hardwired", that is, the tax code is directly written into the code of the program, the analyst can directly modify the tax structure without the intervention of a programmer. Finally, the model can use a wide variety of databases, from DBF formats to Oracle formats and is readily adapted to topics of interest. This type of flexibility in the underlying program structure is critical to the usefulness of microsimulation models.

Output. While any amount of output can be generated from the simulations made using microsimulation models, the most typical include the following tables by income/gross revenue/SIC code group: (1) the distribution of tax burdens under current law, (2) the distribution of tax burdens under proposed law, (3) the "winners" (those with decreases in tax liability) and (4) "losers" (those with increases in tax liability). Summary output includes: total revenue impacts, a forecast of revenue and revenue changes (typically up to 5 years, although it could be longer), and summary statistics by income class (including deductions, taxes paid, total income, etc.). In most cases, the output program is written in the same language as that of the tax calculator.

Extensions of the basic model. One of the newest innovations in microsimulation modeling is the development of more interactive types of models. For example, suppose that a particular tax law change was estimated to reduce corporate tax liabilities by 10 percent. Many people would argue that this could increase investment and output as the price of capital was reduced. If these “feedback” effects occurred, the government could expect greater economic activity, which could *increase* revenues. This cycle could continue-- greater economic activity, greater tax liabilities, higher employment, and so on through the economy. Some microsimulation models have attached these types of macroeconomic feedback effect models so that once the initial impact of the tax law change has been estimated, behavioral changes are introduced, changes to the overall level of income generated in the economy is calculated and distributed among the observations in the sample, and the tax revenues are recalculated. Once again, these types of models rely on very good, detailed information regarding the responsiveness of economic agents to changes in tax policy. These detailed estimates are very difficult to obtain, particularly in transition countries. In Russia, this type of model could be built, but would likely result in revenue estimates that were very sensitive to assumptions about the behavioral changes. However, one should not lose sight of the possibilities of extending a relatively static model to a more dynamic one in the future.

What might a real working model look like? Figure 2 is a more detailed “picture” of the components of a microsimulation model. The basic structure is very similar to that presented above: the data, the tax structure (calculator) and forecasting parameter, the computation, and the report (or output). This figure shows more of the interaction among the components of the model. For example, one can see that the sample data can be stored within the model itself. The flow of the figure shows that the model is interactive in that simulations can be added, changed, amended, and reports can be tailored for particular types of use.

Figure 2: \Prototype of a Microsimulation Model



Moscow City Microsimulation Model

The basic methodology discussed above could be applied to develop a microsimulation model for any type of jurisdiction anywhere in the world. Still, one might question the relevance of these types of models in a transitional economy where the underlying economy and the governmental policies can be changing rapidly. This is precisely the reason for such a model--to be able to capture the implications of these policy changes on the basic level and distribution of revenue. Also, at a time when the private sector is emerging and eclipsing the state sector, new tools must be designed which allow analysis of the policy changes in transitional countries. There is a recognition of the dearth of information on microeconomic responses to changes in the tax system and this can be corrected by the development of an adequate data set and microsimulation model.

One way to conduct microeconomic analysis of firms would be to draw out one or two sample or "representative" firms from an existing database and then to draw inferences about similar firms. This type of analysis is useful and can provide information on the impacts of tax policy changes on certain firms. However, it is difficult to find an "average" firm in each industry. It is also difficult to extrapolate results from any one average firm to the industry as a whole. Finally, given that the number and variety of enterprises is increasing monthly, firm specific analysis would have to be conducted on a continuous basis for an increasing subset of firms, which is a very time-consuming process.

As explained above, the data set for the microsimulation model analysis is crucial to the usefulness of the model. One of the most significant concerns in developing these models in the Russian Federation is that the information necessary would not be available or of sufficient quality to answer questions on the impact on revenues, the distribution of the tax burden and the evolution of the tax base, questions which microsimulation analysis addresses. A majority of territorial tax inspectorates within the City of Moscow are now encoding taxpayer declaration information.

Although there are resource costs involved in collecting microeconomic information, the analysis that can be conducted using microeconomic data more than justifies the cost. Aggregate information does not contain sufficient detail for the mathematical replication of the tax system. For example, let us assume that we have observations on 3 similar enterprises, each of which has the same level of gross profit. However, one enterprise has travel expenses equal to 20 percent of gross profit, one enterprise has travel expenses equal to 10 percent of gross profit and one has no travel expenses. If we change the tax code so that travel expenses are deductible up to 10 percent of gross profit, we would be able to calculate the impact of this deduction on these 3 enterprises. The tax burden would be different for each of these enterprises. However, with aggregated data, the average amount of travel expenses would equal to be 10 percent of gross profit, which would mean that the average tax burden calculated with aggregate data would only be correct for one of the three enterprises. It would not give us information on either what type of enterprises (big or small) or what economic sectors would be mostly affected by this particular measure. While aggregate data can be illustrative of average tendencies, it does not provide sufficient detail to facilitate micro-economic analysis.

The Moscow City Model (MCM) is designed to simulate the impact of changes to the federal tax system and to forecast these revenues. The model focuses on two taxes: the Enterprise Profits

Tax (EPT) and the Value-Added Tax (VAT). This section focuses on the EPT component of the model, although most of the information presented is similar for both components.

The model allows users to forecast revenues and tax liabilities under the current tax system and under various alternative tax structures. Users of the model can easily change the tax parameters to simulate various policy alternatives. Users can directly change:

- tax base variables used in the tax calculators
- tax rates, deductions, credits, and exemption variables
- projected productivity growth in the economy and by industrial sector

The database was developed from true micro-level data. The Moscow City State Tax Inspectorate provided all of the micro-level data from a subset of their territorial tax inspectorates (TTIs). Territorial Tax Inspectorates 27 and 37 provided a sample of microeconomic balance sheet data in 1996 and TTI 33 provided more detailed balance sheet information. The Moscow STI first sanitized the data to ensure taxpayer confidentiality. The data were then provided for analysis and for initial application of the microsimulation models. The results discussed here are based on data from Territorial Tax Inspectorate 33, which provided balance sheet data from 1993 through the 3rd quarter of 1995.

The tax administration unit of any government is the prime candidate for collecting the appropriate data for microsimulation modeling. Currently, Territorial Tax Inspectorates are required to collect and maintain a wide variety of information on taxpayers, ranging form quarterly balance sheet data to enterprise registration information. Most of these data do not have a direct relationship to the primary mission of local tax offices. Given the limited resources of TTIs, one would expect that a prioritization of data would occur, with tax declaration data at the top of the list and other, less pertinent information enter less frequently or without the same amount of screening as the tax declaration information. There is evidence of this prioritization, in that the tax declaration data are becoming more appropriate and cleaner over time which will aid the future development of microsimulation models in the Russian Federation. The following is a detailed discussion of the data used for the MCM.

Enterprises are required to submit quarterly and annual reports on their balance sheets, income statements, and operations to TTIs in conjunction with their tax declarations. These reports form the basis of the current tax administration system, in that cameral audits are conducted on tax declarations, and the balance sheet data are supposed to be used as reference during the cameral audit process. In this phase of the project, data from two reporting forms: (1) the enterprise balance sheet (Form 1) and (2) the financial results of the enterprise (Form 2) from TTI 33 were provided by the Moscow STI to develop the microsimulation model database.

The Form 1 database consisted of 30616 observations, ranging from the 1st quarter of 1993 to the 3rd quarter of 1995. The Form 2 database consisted of 28738 observations, also ranging from the 1st quarter of 1993 to the 3rd quarter of 1995. The time wise distribution of reports is dissimilar, in that the enterprise balance sheet observations are primarily concentrated in the 4th quarter of 1994, while the financial result observations are more evenly spread throughout the sample period (Table 2). In particular, the highest percentage of observations for the Form 1 database is 69.55 percent of the 4^{th} quarter of 1994, compared to 25.14 percent for the Form 2 database which occurs in the 3^{rd} quarter of 1995 (Table 3).

In developing the MCM, the first step of the data analysis was to verify, where possible, the contents of the sample. Verification needs to be the first step of database maintenance to ensure that incorrect data were not entered in the taxpayer database. Verification can also flag enterprise data for closer attention by auditing personnel. In essence, verification simply replicates the reporting structure and checks whether the reported data is mathematically correct. To illustrate the verification process, we examine the properties of the Form 2 database (Table 4).

The financial results database consists of observations of enterprises who have filed tax declarations with TTI 33 and have fulfilled the additional requirement to submit the enterprise balance sheet report and the financial results report. Currently, there is no explicit linkage between the Form 1 and Form 2 data and the tax declaration of the enterprise due to the fact that the Form 1 and Form 2 reports are calculated on an accrual basis while the tax declaration is calculated on a cash basis. However, with the new tax code, which is expected to be approved in 1998, it is expected that the tax declarations will move to an accrual basis.

The verification of Form 2 is split into three components, (1) Financial Results, (2) Uses of Profit, and (3) Payments into the budget. In theory, we should have been able to replicate each component of Form 2 using the reported data. As shown in Box 1, we used the reported information to calculate several variables, including Total Profit, Total Loss, and Net Profit (Loss). We then compared the calculated values with reported values to determine whether the enterprise would pass the verification check. We assigned a tolerance level of 0.01 percent to determine whether the calculated data is consistent with the reported data. The tolerance level was defined as the ratio of the calculated variable to the reported variable. For example, in the case of the first section of Form 2, if 1.0 ⇐ (Calculated Net Profit/Reported Net Profit) ⇐ 0.99, then the calculated data for the specific enterprise was deemed consistent with the reported data.

The possibility existed that inactive enterprises reported information, and we had to account for this occurrence. We defined an inactive enterprise as an enterprise that did not report information on four activities: (1) Proceeds from the realization of goods, works, and services, (2) Proceeds from the realization of other activities, (3) Total Profit or Loss, and (4) a Net Profit or Loss. We could then classify enterprises into three distinct categories, (1) Verified Enterprises, (2) Unverified Enterprises, (3) Inactive Enterprises.

TABLE 2

DISTRIBUTION OF OBSERVATIONS
TTI 33 SAMPLE DATA

Quarter of Observation	Enterprise Balance Sheet Form 1	Financial Results of the Enterprise Form 2
1st Quarter 1993	14	16
2nd Quarter 1993	43	27
3rd Quarter 1993	2327	2024
4th Quarter 1993	2310	2149
1st Quarter 1994	1308	1869
2nd Quarter 1994	873	2146
3rd Quarter 1994	989	3027
4th Quarter 1994	21291	3827
1st Quarter 1995	1455	3346
2nd Quarter 1995	0	3079
3rd Quarter 1995	3	7223
Total Observations	30613	28733

TABLE 3

DISTRIBUTION OF OBSERVATIONS
TTI 33 SAMPLE DATA

Quarter of Observation	Enterprise Balance Sheet Form 1 (percent)	Financial Results of the Enterprise Form 2 (percent)
1st Quarter 1993	0.05	0.06
2nd Quarter 1993	0.14	0.09
3rd Quarter 1993	7.66	7.04
4th Quarter 1993	7.55	7.48
1st Quarter 1994	4.27	6.50
2nd Quarter 1994	2.85	7.47
3rd Quarter 1994	3.23	10.53
4th Quarter 1994	69.55	13.32
1st Quarter 1995	4.75	11.65
2nd Quarter 1995	0.00	10.72
3rd Quarter 1995	0.01	25.14

TABLE 4

VERIFICATION OF FINANCIAL RESULTS SECTION
RAW DATA
FORM 2

Quarter	Number of Observations	Verified Observations	Unverified Observations	Inactive Enterprises
1st Quarter 1993	16	15	0	1
2nd Quarter 1993	27	25	1	1
3rd Quarter 1993	2024	1825	131	54
4th Quarter 1993	2149	2018	55	37
1st Quarter 1994	1869	1747	45	57
2nd Quarter 1994	2146	1853	37	32
3rd Quarter 1994	3027	1858	54	18
4th Quarter 1994	3827	1952	50	35
1st Quarter 1995	3346	1640	64	37
2nd Quarter 1995	3079	1846	41	23
3rd Quarter 1995	7223	1858	47	146
Totals	28733	16637	525	441

Box 1

Verification of Financial Results Statement
Financial Results Section

Calculated Reported Profit or Loss

=	S90N (Profit)	or	S90K (Loss)
=	Total Profit	-	Total Losses
=	S80N	-	S80K

Calculated Total Profit - Total Loss

=	S80N (Profits) - S80K (Losses)		
=	Proceeds from realization of goods	=	S010N
	- Value Added Tax	=	S015K
	- Net Excise Tax Payments	=	S030N - S015K - S030K
	- Manufacturing Expenses	=	S040K
	+ Net Results of Other Realizations	=	S60N - S60K
	+ Net Non-Realization Transactions	=	S070N - S070K

= S101N - S015K + (S030N - S015K - S030K) - S040K + (S60N - S60K)
+ (S70N - S70K)

Our initial analysis found that approximately 94 percent of all the observations for the Financial Results section of the Financial Results Statement (Form 1) could be verified. However, when we attempted to verify variables in the following two sections of the Financial Results Statement, we did not achieve the same level of success as with the first section of the reporting form. We first cleaned the data to delete duplicate observations and combine multiple observations for the same enterprise. We then designed a series of verification algorithms to test the validity of the sample data past the first section. For example, one of the verification algorithms tested whether calculated payments to the budget were within 0.01 of reported payments to the budget. The objective was to test the consistency of the data across sections of the report.

The first test we conducted on the processed data was to determine whether the reported profit (loss) was within 0.01 of the sum of the uses of profit, which was the second section of the Financial Results Statement. We were only able to verify payments to the budget in approximately 19.33 percent of the enterprises across the sample (Tables 5 and 6). This ranged from a high of 34.88 percent in the 4th quarter of 1993 to a low of 9.07 percent in the 3rd quarter of 1994. This presents an obvious concern in that if tax administrations are not able to verify payments to the budget by specific enterprises, then they are faced with an improbable task of assigning discrepancies in budgetary payments.

TABLE 5

VERIFICATION OF FINANCIAL RESULTS STATEMENT
PAYMENTS TO BUDGET AND VALUE ADDED TAX VERIFICATION

Quarter	Number of Enterprises	Verified Net Profits	Verified Budgetary Payments	Verified VAT Payments
1st Quarter 1993	16	15	3	3
2nd Quarter 1993	27	25	4	5
3rd Quarter 1993	2010	1825	556	248
4th Quarter 1993	2110	2018	736	267
1st Quarter 1994	1849	1747	362	398
2nd Quarter 1994	1922	1853	227	330
3rd Quarter 1994	1930	1858	175	299
4th Quarter 1994	2037	1952	582	291
1st Quarter 1995	1741	1640	288	409
2nd Quarter 1995	1910	1846	260	411
3rd Quarter 1995	2051	1858	355	494
Totals	17603	16637	3548	3155

TABLE 6

VERIFICATION OF FINANCIAL RESULTS STATEMENT
PAYMENTS TO BUDGET AND VALUE ADDED TAX VERIFICATION

Quarter	Number of Enterprises	Verified Net Profits (percent of total)	Verified Budgetary Payments (percent of total)	Verified VAT Payments (percent of total)
1st Quarter 1993	16	93.75	18.75	18.75
2nd Quarter 1993	27	92.59	14.81	18.52
3rd Quarter 1993	2010	90.80	27.66	12.34
4th Quarter 1993	2110	95.64	34.88	12.65
1st Quarter 1994	1849	94.48	19.58	21.53
2nd Quarter 1994	1922	96.41	11.81	17.17
3rd Quarter 1994	1930	96.27	9.07	15.49
4th Quarter 1994	2037	95.83	28.57	14.29
1st Quarter 1995	1741	94.20	16.54	23.49
2nd Quarter 1995	1910	96.65	13.61	21.52
3rd Quarter 1995	2051	90.59	17.31	24.09
Average		94.29	19.33	18.17

The second test we conducted on the processed data was to determine whether the reported VAT payment in the first section of the Financial Results Statement matched the reported VAT payment or liability in the third section of the Financial Results Statement. In essence, we were checking if the Financial Result Statements were internally consistent. Once again, the promising results of the verification process of the first section were erased by the low quality of the data. We were only able to verify VAT payments or liability for approximately 18.17 percent of the enterprises in the sample (Tables 5 and 6). This ranged from a high of 24.09 percent in the 3rd quarter of 1995 to a low of 14.29 percent in the 4th quarter of 1994. The concern is that the Financial Result Statements are not internally consistent in the vast majority of cases.

These verification exercises are extremely important and attest to the complications faced by analysts in developing microsimulation models. However, the continued data development for the MCM will continue to improve the accuracy and usefulness of the model.

The tax calculator is written for "current law" that is, the law that was in effect in 1996. Changes to that law can very easily be made through a user-friendly menu which allows the user to change tax rates, deductions, brackets, etc. The model operates through Windows and uses a Borland Delphi interface. The entire model can be run on a 486 or higher personal computer.

By activating the current law calculator of the model (which is done using the menu system), the user can specify tables of tax liability by individual firms in the sample, by industry group, by asset level, or by some other identifier. This is extremely important because not only is it useful to know the static revenue estimate of a tax law change, it is also important to know which firms are the "winners" (reduced tax liability) and which are the "losers" (increased tax liability). This is true because at some points in time, the country may wish to equalize treatment among types of firms, or the country may want to provide preferential treatment for some industries. Without a microsimulation model, it is very difficult to find out which firms win and which lose.

The microsimulation model for the Enterprise Profits Tax consists of 8 modules. The 8 modules are primarily interfaces designed to help the analyst define a base and alternative tax structure, input exogenous economic parameters, and calculate the tax liability for each taxpayer in the sample. Unlike many previous generations of microsimulation models, the Enterprise Profits Tax model is parametrically driven, in that the parameters that define the structure of can be defined at run time. This means that the analyst can incorporate changes to the tax structure into the microsimulation analysis without having to define an external parameter file or change the code of the program.

As explained above, at the heart of the microsimulation program is the tax calculation routine. The tax calculator takes the parameters defined by the analyst, such as gross revenues, exemptions, deductions, and tax rates, and passes the microeconomic data through a mathematical representation of the tax structure to calculate the users tax liability. In order to accurately represent the tax structure, microeconomic information must be of sufficient detail so as to replicate the taxpayer's tax declaration.

After the tax liability has been calculated for each individual taxpayer in the sample, exogenous economics factors may be applied to the sample to construct a microeconomic forecast.

The sample is also calibrated to be as representative as possible to the actual distribution of taxpayers and revenue collections for the geographical. A series of reports may then be generated by the analyst to illustrate the impact of the proposed change to the tax system. The ability to simulate and forecast results for the entire universe of taxpayers depends on the quality and availability of data. Even the most sophisticated microsimulation models cannot correct for the lack of desegregated microeconomic data.

One of the initial concerns during the implementation phase of the project was the availability of data of sufficient quality and quantity to conduct statistically valid microsimulation analysis. As previously discussed in order to construct a mathematical representation of the tax structure, it is necessary to have access to tax declaration data. Tax declaration data, in combination with the financial statement and enterprise balance sheet reports of the enterprise, allows the analyst to recreate the majority of the calculations used to prepare the tax declaration for submittal to the tax inspectorate. As explained above, since true tax declaration are not widely available, the MCM was built using two panel databases contained information on the balance sheet and the financial results statement of legal entities.

How does the actual Moscow City microsimulation model work? As already noted, the microsimulation model for the Enterprise Profits Tax has 8 modules, ranging from a simple description of the program to the tax calculator. The start of the microsimulation process is the choice of a microeconomic database. The microeconomic database can consist of a variety of possible data sources, however, for the best possible operation of a microsimulation model, taxpayer declaration data should be used. In the case of the first phase of the project, taxpayer declaration data was not available, so microeconomic data on the financial results statement and the enterprise balance sheet was used. From the microeconomic database, a sample of information is selected by the analyst and "cleaned." The cleaning process can delete erroneous observations from the database. For example, inactive enterprises can be deleted from the microeconomic database to reduce the size of the database and to speed processing time. The resulting database or databases represent inputs into the microsimulation model.

After the data source is specified using the model, the data record is passed to the microsimulation model and held in memory until needed. One should note that there is a difference between observations and records. A data observation is a specific, possibly multidimensional, point of information that contains information on gross revenues, deductions, exemptions, tax liabilities, and other pieces of information that are contained in the microeconomic database. A data record is the structure of the observations, that is, what types of variables are contained where in the database. A data record is the structure of the observations, that is, what types of variables are contained where in the database. The microsimulation model retains the data record in memory until it is necessary to read the data observations.

The data record serves another purpose in that it is used by the analyst to define the default and alternative tax structures. The data record contains not only the variable positions and types but also the variable names. Since the analyst has constructed the microeconomic database, it is safe to assume that the analyst is familiar with the variables in the database.

The analyst would choose the variables that are used to calculate gross revenues from the available database. These variables would be stored in a parameter file that could be immediately used by the analyst in the microsimulation model or saved for analysis at a later date. The analyst would continue to select variables for gross revenues, deductions, exemptions, the tax rates and brackets, and industrial identifiers. All these variables would be saved to an external parameter file within a parameter record. Note that the program does not save actual microeconomic data but only a record of which variables are used to calculate which categories.

Another component of the parameter file are the exogenous economic parameters that are used to either calibrate the microeconomic database to represent the population of taxpayers, construct a microeconomic forecast, or both. The analyst can calibrate the sample to actual collections or the number of active taxpayers within a specific geographical location. The analyst can also enter in an inflation rate forecast, growth rates of Gross Domestic Product for the entire Moscow region, or by specific industry, and growth by number of firms by specific industry. However, unlike the record written out in the definition of the default or alternative file structure, the actual variable values are written out to the parameter file at the end of this module. The set of exogenous economic parameters is considered unique to each parameter file and is therefore saved within each parameter file.

After the analyst has specified a default and alternative parameter file, the microsimulation module can be run to calculate the default and alternative tax liabilities for each observation in the selected database. This is the strength of microsimulation models with respect to aggregated simulation models. Microsimulation models investigate the impact of changes in the tax structure on the tax liability of individual taxpayers. Aggregate models, while useful, can not be used to draw inferences as to the microeconomic impact of changes to the tax system. The resulting output of the microsimulation model are two sets of tax liabilities, calibrated to the set of exogenous parameters. Also included in the output are calculated gross revenues, deductions, and exemptions, so that the analyst can directly interpret how the changes in the tax structure influenced the calculation of each taxpayers' liability.

After the microsimulation module has been run, the analyst can choose from a variety of pre-formatted tables that report on the impact of the changes between the default and alternative tax structures. Of course, if these pre-formatted reported do not suit the needs of the analyst, then they can use the *ReportSmith* component of *Delphi* to create their own reports. The analyst can repeat the microsimulation process at any time, including the specification of new data and parameter files.

Results

The MCM development and use has provided a number of "results" -- some direct policy simulations and some externalities from the development of the model. First, the analysis of the microeconomic data provided from the two subordinate Territorial Tax Inspectorates illustrated the need for tax administration reform, as the data requirements for the Russian tax system were only increasing over time. Second, the Moscow STI needed to create a system by which microeconomic data was forwarded from Territorial Tax Inspectorates to the Central Office of the Moscow STI. Third, the current technological capability of the Central Office of the Moscow STI needed to be upgraded to support the database requirements of microsimulation analysis. Fourth, the initial microsimulation models needed to be revised to incorporate changes to the tax system and proposed changes to the tax code. These are important lessons in the model development.

Regarding the policy simulations for which the MCM was developed, let us consider a particular tax policy change, and follow through how a microsimulation model can provide valuable input into the policy dialogue. The current MCM contains data for one tax inspectorate (TTI 33). The baseline (current law) corporate profits tax liability for the firms in that inspectorate is 5.096 million rubles for the sample observations. Using the model, we can answer the following policy-related question: What is the impact of increasing the tax rate from 32 percent to 33 percent and add a one time increase of fixed capital depreciation of 20 percent?

We can specify this rate change and change in depreciation rules using the menu system of the MCM. Once the tax law change has been made, we produce the attached output tables. Table 7 is a simple table which shows the total level of change in tax revenue associated with the proposed tax law. The change is a reduction in revenue of 0.38 million rules, less than percent of current law liability. This does not appear to be an important tax change, if the total change in revenue is such a small a percent. However, we do not know whether all firms realize a reduction in tax liability, or if some firms realize a reduction while others see an increase, and whether certain sectors bear the largest reduction.

Table 8 provides much more detailed information about the distribution of the tax law change. As seen there, it is the firms with relatively larger holdings of assets that realize the biggest reductions in tax liability, while the firms with relatively small asset holdings see an increase in their tax liabilities. This might surprise policy makers--a law which decreases tax liability in aggregate, increases the tax liability of many smaller firms, while providing relief to larger firms (measured in terms of asset size). This may or may not be an intent of the proposed law, but without this model, policy makers are forced to make decisions with much weaker information.

Table 9 provides another look at the distribution of the tax change. From Table 9 it becomes obvious that the entire tax decrease is afforded to one firm, in one industry. The distribution of this reduction may not be an intention of the policy at all--or if the policy was aimed at one industry, the microsimulation model provides evidence that the targeted relief was successful.

This one example provides one illustration of the MCM. This microsimulation model is built to inform the policy debate on virtually any issue dealing with changes in tax structure: rates, exemptions, deductions, etc. As evidenced by the preceding example, the MCM provides more than just an answer to the revenue cost of a particular proposal. The model is designed to provide information also on which firms are winners, which losers, what types of industries are most affected, etc. This additional information is critical to tax policy debates and this information is best provided by microsimulation models.

TABLE 7

SUMMARY REPORT
Corporate Profits Tax

Number of Firms	Default Tax Revenue	Alternative Tax Review	Net change
584	5,096,601	5,057,639	-38,963

Simulation of Territorial Tax Inspectorate 33, 1994 2nd quarter.
Change Tax Rate from 32 percent to 33 percent and added one time increase of Fixed Capital Depreciation of 20 percent.

TABLE 8

ENTERPRISE ASSET REPORT
Corporate Profits Tax

Asset	Range	Existing Firms	New Firms	Default Revenue	New Revenue	Difference	Winners	Losers	Unchanged
1	99	432	432	2,650,746	2,733,582	82,836	0	172	260
100	999	4	4	44	45	1	0	0	4
1,000	9,999	31	31	394	399	5	1	0	30
10,000	49,999	45	45	57,519	58,158	640	2	11	32
50,000	99,999	34	34	37,488	37,555	67	5	12	17
100,000	999,999	11	11	68,348	68,337	-11	2	6	3
1,000,000	4,999,999	19	19	292,842	278,866	-13,976	5	6	8
5,000,000	9,999,999	5	5	237,595	205,167	-32,428	3	1	1
10,000,000	49,999,999	0	0	0	0	0	0	0	0
50,000,000	99,999,999	3	3	1,751,627	1,675,531	-76,095	1	1	1

Simulation of Territorial Tax Inspectorate 33, 1994 2nd quarter.
Change Tax Rate from 32 percent to 33 percent and added one time increase of Fixed Capital Depreciation of 20 percent.

TABLE 9

ECONOMIC SECTORS REPORT
Corporate Profits Tax

Sectors	Firms (before)	Firms (after)	Tax (before)	Tax (after)	Net Changes	Winners	Losers	Even
Industry Group 1	4	4	1,852,429	1,746,665	-105,764	2	1	1
Industry Group 2	259	259	931,552	935,524	3,972	10	89	160
Industry Group 3	19	19	456,488	470,573	14,085	0	7	12
Industry Group 4	2	2	6,154	6,346	192	0	1	1
Industry Group 5	58	58	153,180	157,961	4,782	0	21	37
Industry Group 6	0	0	0	0	0	0	0	0
Industry Group 7	0	0	0	0	0	0	0	0
Industry Group 8	242	242	1,696,799	1,740,569	43,770	7	90	145

Simulation of Territorial Tax Inspectorate 33, 1994 2nd quarter.
Change Tax Rate from 32 percent to 33 percent and added one time increase of Fixed Capital Depreciation of 20 percent.

Conclusions and Future Research

It is obvious that microsimulation models are crucially important to the tax policy debate in Russia and in countries around the world. The models provide more detailed information regarding the impacts of tax law changes than any other technique currently available in Russia. The use of these models worldwide attests to their importance and usefulness.

The hindrance to using microsimulation models in Russia is a lack of appropriate data and a lack of historical information on the impacts of tax law changes on behavior of firms and individuals. The first data problem can and is being addressed. Data development is occurring within the STS, and within some oblast finance departments. The second issue will be remain for the near future. Within the next five years, however, Russia should look to extending it microsimulation modeling techniques to include more dynamic estimates.

Bibliography

Bahl, Roy, Richard Hawkins, Robert E. Moore, and David Sjoquist, (1993), "Using Microsimulation Models for Revenue Forecasting in Developing Countries," in *Public Budgeting and Financial Management*, vol. 5, no. 1: 159-186.

Bahl, Roy and Sally Wallace, (1994), "Consultation on Philippine Tax Reform," Report prepared for USAID-Manila, August 10, 1994.

Baekgaard, Hans (1996), "A Microsimulation Approach to the Demand for Day Care for Children in Denmark," in A. Harding, ed., *Microsimulation and Public Policy* (Amsterdam: Elsevier Press).

Bergmann, Barbara R. (1975), "Combining Microsimulation and Regression: A Reply," in *Econometrica*, vol. 43: 529-531.

Kapur, Vishnu, Anil Grupta, and Tom McGirr, (1997), "Microsimulation and Sales Tax Reform in Canada," in *Microsimulation in Government Policy and Forecasting: International Conference on Combinatories, Information Theory and Statistics* (Portland, ME: University of South Maine, Forum for Interdisciplinary Mathematics).

Klevmarken, Anders and Paul Olovsson (1996), "Direct and Behavioral Effects of Income Tax Changes-Simulations with the Swedish Model MICROHUS," in A. Harding, ed., *Microsimulation and Public Policy* (Amsterdam: Elsevier Press).

Mirer, Thad W., and Jon K. Peck (1975), "Combining Microsimulation and Regression: A Comment," in *Econometrica*, vol. 43: 523-528.

Orcutt, Guy H. (1960), "Simulation of Economic Systems," in *American Economic Review*, Vol. 50: 893-907.

Orcutt, Guy H. (1957), "A New Type of Socio-Economic System," in *Review of Economics and Statistics*, vol. 58: 773-797.

Orcutt, Guy H., M. Greenberger, M. Korbal, and Alice Rivlin, (1961), *Microanalysis of Socioeconomic Systems: A Simulation Study* (New York: Harper and Row).

Policy Research Center, Georgia State University, (1996), *Microsimulation Model User's Manual*, (Atlanta: Georgia State University).

Policy Research Center, Georgia State University, (1997), "Final Report: Fiscal Management Project, Phase I," (Atlanta: Georgia State University).

Symons, Elizabeth and Neil Warren (1996), "Modelling Consumer Behavioral Response to Commodity Tax Reforms," in A. Harding, ed., *Microsimulation and Public Policy* (Amsterdam: Elsevier Press).

U.S. Social Security Administration, Wixon, Bernard, Benjamin Bridges, and David Pattison (1987), "Policy Analysis through Microsimulation: The STATS Model," in *Social Security Bulletin*, vol. 50: 4-12.