

LATAX

A multi-country flexible tax micro-simulation model

Institute for Fiscal Studies

Laura Abramovsky and David Phillips

September 2015

LATAX is a multi-country flexible tax micro-simulation model developed by researchers from the Institute for Fiscal Studies (IFS) for the analysis of VAT, excise duties, income tax and social security contributions, as well as (non means-tested) price subsidies. It is designed to be a simple public tool for micro-analyses of future or counterfactual tax reforms in Mexico, El Salvador, and potentially other countries. LATAX can quantify the revenue and distributional impact of tax reforms under both the assumption that individuals do not change their behavior as a consequence of changes in taxes, and the assumption that individuals react to these changes along specific margins.

LATAX is a user-friendly simulator program designed so that users do not need to edit the main program code but can instead make changes in i) tax system parameters and ii) estimate accounting for behavioral impacts on households income/expenditure. Based on the available data and the user-defined tax parameters, the program produces indirect tax payments, the direct tax base, and direct tax payments before calculating the revenue effects of the reforms and the impact of the tax changes across the income / expenditure distributions and by household types. Full details of the simulator program can be found in the background papers by Abramovsky et al (2010, 2011 and 2012).

LATAX is based on IFS micro-simulation model for Mexico (MEXTAX). LATAX draws both upon IFS experts' experience in analyzing tax and benefit reforms in the UK, and on the experience and methods used by other researchers in Mexico and other Latin American countries. MEXTAX is now included/built-in in LATAX.

This user manual explains how to use LATAX for analysis in **Mexico**¹ and includes:

- (1) Introduction – what is LATAX?
- (2) Overview of program structure
- (3) Installation requirements and procedures
- (4) Running the model: a worked example
- (5) Interpreting outputs of model: distribution, revenue, behavior, etc.
- (6) Constructing the input data
- (7) Appendices which provide details on: the reforms modelled (appendix A); full details on the user-editable interface commands and parameters (appendix B); information on the variables in the input and output datasets (appendix C); and information on how we estimated a consumer demand system for including in LATAX.

¹ Most of this manual, with the exception of explanations of the consumer demand module, and the creation of the input datasets is applicable to use of the LATAX model for El Salvador.

It is written using a ‘worked example’, the 2010 Mexican reforms, with screen shots to help make it more ‘concrete’ and to provide a practical example to help people with.

1 Introduction

The LATAX model is a flexible static microsimulation model developed in STATA software, which has been designed by IFS researchers to be a public tool for analyses of future tax reforms.²

LATAX can quantify the revenue and distributional impact of tax reforms (it calculates taxes paid for each household, and produces summary distributional and revenue analysis) under both the assumption that individuals do not change their behavior as a consequence of changes in taxes, and the assumption that individuals react to these changes along specific margins:

- i) Labour supply (if suitable elasticities of participation and hours of work are available);
- ii) VAT pass-through - the degree to which indirect taxes are passed on to consumers by producers (if suitable estimates of VAT pass-through are available)
- iii) Consumer demand - consumers’ responses to changes in prices induced by changes in taxes (using estimated price elasticities in the case of Mexico).

LATAX examines the distributional and revenue impacts of tax reforms at one point in time. It accurately models rules and structures of complex tax systems. It covers:

- i) Income tax³
- ii) Value Added Taxes
- iii) Excise duties
- iv) Social security contributions

LATAX models the tax system at the level of the individual (or household) using survey or micro-level administrative data; it requires the use of detailed household level data with detailed information on demographics, income and expenditure (the precise requirements are discussed in Section 6). Sample weights allow one to conduct distributional analysis for the entire population and calculate aggregate revenue effects of tax reforms using the sample of individuals/households for which data is available.

LATAX is based on a modular approach:

- User amends “interface” module which contains settings on the type of analysis to be performed and certain assumptions to be made
- And the “parameter” modules which include tax rates, thresholds etc. under the base and reform systems – it can handle one base system and two reform systems
- Tax calculations and distributional analysis modules are not ‘system specific’ and need not be amended for a large number of reforms
- This means it is easy to do many actual and counterfactual reforms with limited knowledge of how the program calculates taxes

² A tax micro-simulation model can be static or dynamic, and can include no-behavioral response, assumed behavioral response or estimated behavioral response.

³ In existing applications of LATAX for Mexico, only income taxes on labour income have been simulated due to low quality information capital income in the household survey used.

It outputs a household-level dataset with incomes, expenditures and tax payments and log files with summary distributional and revenue figures.

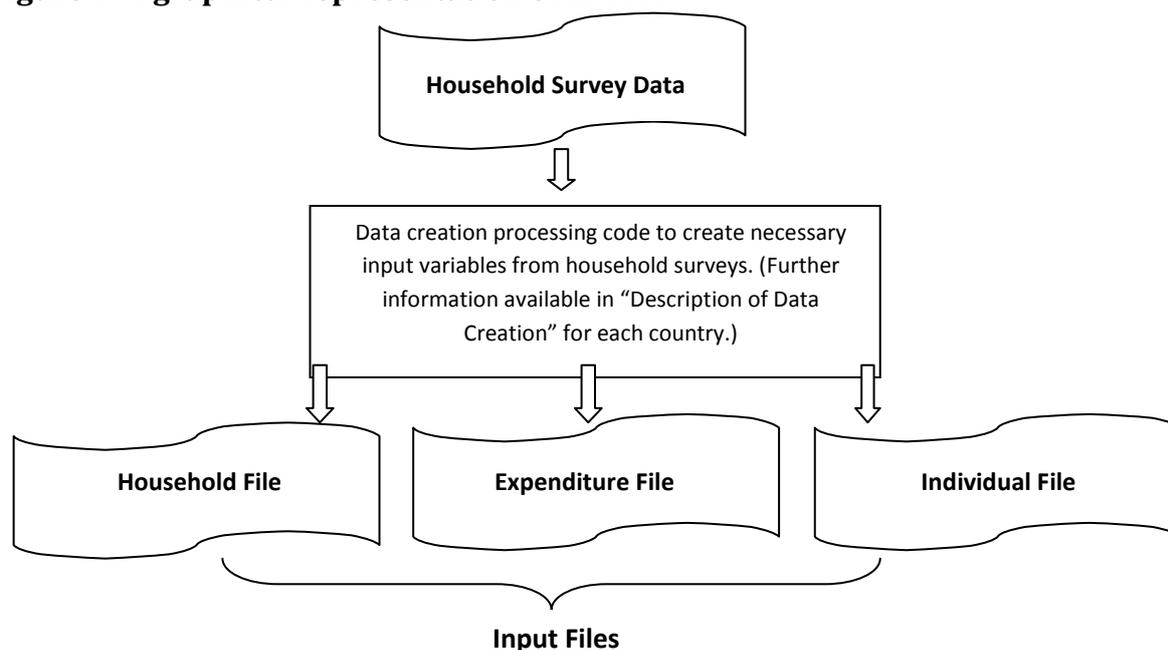
It is hoped that the model will be useful to academics, civil society organisations, international bodies such as the World Bank, and civil servants. In order to help such users get the most out of LATAX, this manual explains how to set up and use the model, using a worked example of simulating the distributional and revenue impacts of the 2010 Mexican tax reforms. Details of the reforms being analysed can be found in Appendix A.

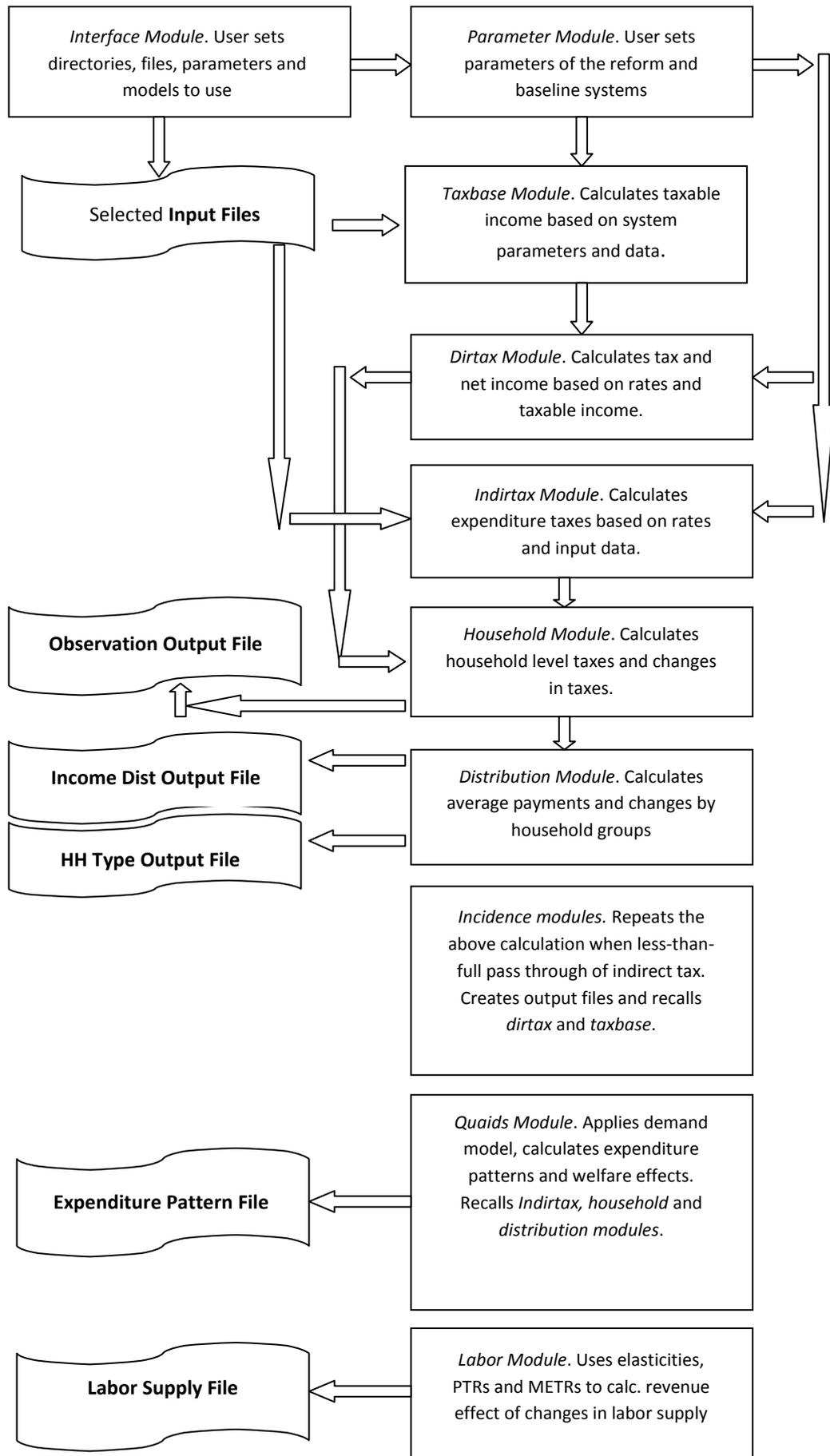
2 Overview of program structure

LATAX is written in Stata code and is designed so that users do not need to edit the main simulation code. Instead, basic instructions such as the names of input and output datasets, which tax systems to use, and whether to run behavioral response modules can be input either interactively via the Stata command line or input into a main *interface* module, and the parameters of the baseline and reform tax systems changed by amending system *parameters* modules (which define the basic structure and tax rates).

Based on the input data and the user-defined tax parameters, separate modules then calculate indirect tax payments, the direct tax base, and direct tax payments before calculating the revenue effects of the reforms and the impact of the tax changes across the income / expenditure distributions and by household types. Separate modules can be turned on and off according to need to allow for less-than-full pass-through of changes in indirect taxes to changes in consumer prices, as well as to model labor supply (or more correctly, taxable income) and consumer demand responses to tax changes (if a demand system has been estimated for the country LATAX is being used for, which is presently only the case for Mexico). It has been designed in this way so that users do not have to edit the main program code even if they wish to make fairly major changes to the tax system (e.g. introducing additional tax rates) or the input data (such as additional sources of income or expenditure categories). Figure 1 shows the basic structure of the program. We then describe each module in turn.

Figure 1. A graphical representation of LATAX





main_module.do

This interface main module contains user-edited instructions on: the directories in which the simulation code, input and output data, and parameter code can be found; the name and location of logfiles; the names of input and output datasets; the existing and reform systems to be used in the analysis (sys1, sys2, sys3); and runs the tax simulator. There are options for whether one wants to conduct the behavioral analysis. A worked example explaining how to use the *main_module* can be found in section 4, whilst full information on all the options can be found in Appendix B.

loadprogs.do

This module calls the programs that perform the tax, revenue and distributional calculations and that estimate the behavioral response to a particular set of reforms (given the assumptions provided to LATAX). This is done as follows:

- First, for each of the three tax systems chosen in *main_module.do* (sys1, sys2, sys3), load the parameters file (*params_X*) from the correct parameters folder, and then call *calc_indirtax.do*, *set_taxbase.do* and *calc_dirtax.do*.
- Then, call *do_household.do* and *do_distanalysis.do*
- Finally, if selected (using *consmo*, *indir* and *labmod* as set in *main_module.do*) call *quaid.do*, *indirect_incidence.do* and *labor.do*.

params_X.do

The user defines the parameters of the VAT, duties, income tax (ISR) and Social Security tax systems in the parameter do files. When naming the do files they should always be of the form *params_X.do*, where X is the *numerical* name of the particular system (and corresponds to sys1, sys2 or sys3 as defined in *main_module.do*). Existing parameters modules should be used as templates and it is important that the names of scalars are not changed (although of course the values can be). A worked example explaining how to amend the *params_X* modules can be found in section 4, whilst full information on all the parameters can be found in Appendix B. Note that the descriptions of the calculation modules, below, refer to particular parameters of the tax systems defined in Appendix B.

calc_indirtax.do

This program first uses the VAT and duties rates that applied at the time of the survey underlying the expenditure input data (vectors OVAT, OBVAT and ODUTIES) to calculate pre-tax prices. Total pre-tax expenditure is also calculated. Next, the VAT and IEPS rates that apply under the system under consideration (vectors VAT, BVAT and DUTIES) are used to calculate the amount of VAT and IEPS due under that system. The calculation is done by good and then summed over goods to give a total per household.

set_taxbase.do

First, depending on the value of the global *indic_incidence*, defined in the *main_module*, either the standard input data is loaded (*indic_incidence* = 1) or the data adjusted for less than-full pass-through (*indic_incidence* = 2 or 3). Then using the amount of each income from each source and the exemptions (EXEMPT, PEXEMPT, MPEXEMPT, SSEXEMPT, SSPEXEMPT, and SSMPEXEMPT) the taxbase for income tax and social security contributions is calculated.

Sections allowing for deductions of certain expenses (such as medical insurance) are currently commented out as the inability to model income tax on capital and self employment income (partly due to poor data) means that one cannot calculate the maximum amount deductible (which is typically a fraction of taxable income).

calc_dirtax.do

The first part of the program calculates social security contributions using the social security bands and rates (LSOCSECX_Y, SOCSECX_Y) to calculate fixed quotas for social security contributions: that is the amount of contributions paid on income up to the start of each band. This means that when calculating the amount of social security contributions paid, this can be added on to the amount paid on income within the band an individual finds themselves in, simplifying calculations considerably. The same process is then done for income tax payments, after deducting the amount of social security contributions paid from the income tax base if DEDUCTSOCSEC==1.

The amount of subsidy for employment income is calculated using the tax base for income tax and the ACRED and LCRED amounts defined for the system.

Once income tax and social security payments are calculated, net income is calculated as (gross income) – (income tax) – (social security contributions).

Depending on the value of the global indic_incidence, the output is saved either as a standard file (indic_incidence =1) or as output data for the analysis of less-than-full pass-through (indic_incidence =2 or 3).

do_household.do

This module first loads the relevant variables from the input data (individual and household files) and then merges in the results of the indirect and direct tax calculations for the base system (sys1) and the two reform systems (sys2 and sys3). Individual and household level changes in tax payments under the reform systems are then calculated.

Household net monetary and non-monetary income is calculated from the tax simulation outputs if netincometype==1 in the interface main module.

Individual-level and household level output files are saved.

do_distanalysis.do

This program first calculates the revenue effects of each reform by summing (and appropriately grossing using sample weights) the changes in tax payments for each tax. This is saved in a log file.

The program then calculates equivalised incomes for each equivalence scale (100/100/100, 100/80/50 and 100/50/30) and equivalised income and expenditure decile groups.

The global disttype (defined in *main_module.do*) tells this module whether the full set of distributional analysis should be conducted (disttype=1) or only a partial analysis (disttype=2).

Average cash gains/losses per household by decile group are calculated as (minus) the sum of the change in tax payments in each particular decile group, divided by the number of households in each particular decile group. The average proportional gains/losses per household by decile group are calculated as (minus) the sum of the change in tax payments in each particular decile group, divided by the total income/expenditure of households in each

particular decile group. In addition, the proportion of the increase (or decrease) in revenue attributable to each decile group is also calculated for each tax and in total, and compared to the proportion of total income/expenditure attributable to each decile group.

The results are displayed in tables which are saved in log files.

The process is repeated for gains/losses by household type.

indirect_incidence.do

This program allows one to carry out analyses assuming that indirect taxes are partly incident on wages and profits (instead of fully on prices).

The program first loads output data from the indirect tax calculations (performed in *calc_indirtax.do*) and input data from the files *indata_ad* and *indata_hh*. Calculations then proceed as follows.

First, post tax-reform consumer prices under the assumption of less-than-full VAT and duties pass through, and the associated direct effect of the change in VAT and duties on consumers are calculated using the assumed pass-through rate. Then new consumer prices are used to calculate pre-tax prices and the amounts of VAT and duties paid under the reform systems.

The next stage is to allocate the part of the burden not borne directly to consumers to gross formal sector employment income and (net) capital income according to the ratio assumed in *main_module.do*. It is assumed that the amount of income of each individual with these sources is reduced by the same proportion.

We recalculate the amount of tax paid and net income given the changed gross income calculated in the previous stage by recalling *set_taxbase* and *calc_dirtax*. When doing this we change the value of *indic_incidence* to 2 so that *set_taxbase* and *calc_dirtax* know that they are being called by *indirect_incidence* rather than *load_progs*.

The module then performs the same functions as *do_household.do* and *do_distanalysis.do*, outputting to log files and an individual-level data file (*logfiledecile_ind*, *logfilehhtype_ind*, *logfilerevenue_ind* and *outdata_ad_ind*). It should be noted that the tables in the decile and household log files show the proportional changes in net income and expenditure (assuming fixed purchase quantities) rather than changes in taxes paid. This is because when pass-through is less than complete, the changes in taxes paid and the gain/loss to households differs.

labor.do

This program allows one to estimate how taxable employment income responds to changes in both direct and indirect taxes, given an assumed set of elasticities (on the intensive and extensive margin).

First, the household-level average tax rates on expenditure are calculated, accounting for informal transactions. These are then added to average income tax and social security rates to get the participation tax rates (PTRs), and to marginal income tax and social security rates to get marginal effective tax rates (METRs).

The user then defines the hours elasticities (intensive margin) and the participation elasticities (extensive margin), which may vary by demographic group. The changes in taxable employment income and the associated changes in tax revenues are then calculated using the calculated PTRs and METRs and the assumed elasticities.

Finally, demographic variables are created and the changes in taxable labor income and revenue are outputted in `logfilelabor`.

quaid.do

This module allows one to estimate how consumer spending patterns change following tax changes, and how such changes in spending patterns affect the revenues from the tax changes, and consumer welfare. It makes use of demand system estimates calculated in `quaid`_estimation.do (see Appendix D).

First, the expenditure on each of the 12 demand system goods categories is calculated using the NUMGOODS categories of the MEXTAX indirect tax simulator. This is done to calculate total expenditure on goods included in the demand system, which together with prices (read in from `indata_prices`) and the coefficients from the demand system (read in from `indata_demand`), is used to calculate indirect utility and the estimated shares for each of the 12 demand system goods categories for each household.

The price effects of tax reforms are then calculated and are used to estimate the shares for each of the 12 goods categories following the reforms. The expenditure function is then calculated to calculate the compensating variation (CV) for each reform, which (after adding on the change in taxes paid on goods which we cannot include in our demand system) is our measure of the welfare effect of the tax reforms accounting for substitution possibilities. We also calculate the change in expenditure required to obtain the same level of utility if substitution were not possible so that we can evaluate the extent to which substitution possibilities ameliorate the impact of tax changes.

The changes in consumer welfare by decile group and household type are calculated and outputted in `logfiledemand_welfare`; the changes in expenditure shares are outputted in `logfiledemand_spending`; and the changes in revenues are outputted in `logfiledemand_revenue`.

3 Installation requirements and procedures

LATAX is written in Stata and has been successfully used with versions 10, 11 and 12. A key requirement for using the program is therefore to have Stata installed (contact your system administrator or see www.stata.com for information on how to obtain and install Stata if you do not already have access to it).

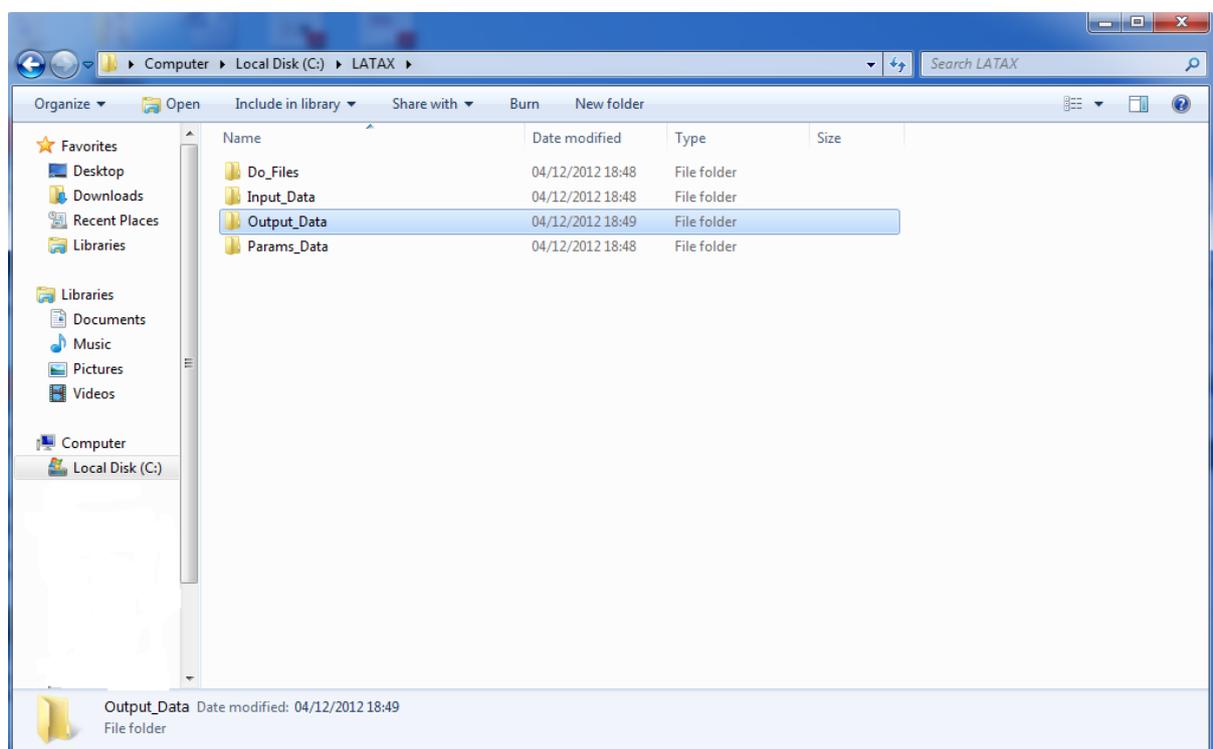
With Stata installed, download the LATAX package from the [LATAX website](#). The downloaded file is a zip file: the do files making up the LATAX program need to be extracted and placed in folders from they can subsequently be called when using the program. There are two types of do files: program modules (`main_module.do`, `loadprogs.do`, `calc_indirtax.do`, `set_taxbase.do`, `calc_dirtax.do`, `do_household.do`, `do_distanalysis.do`, `indirect_incidence.do`, `labor.do` and `quaid`.do) and 3 example parameters modules which are used in the worked example in this user manual and can be used as the basis for your parameters modules (`params_2008`, `params_2010` `params_20101`). Where the different types of files are placed is, to a large extent, up to you: the program will work if they are all placed in one folder, or the program files and parameters files are placed in separate folders. However, it is recommended, initially, to use the following convention:

- Create a folder called “LATAX”

- Create a sub-folder in “LATAx” called “Do_Files” and place the program modules (*main_module.do*, *loadprogs.do*, *calc_indirtax.do*, *set_taxbase.do*, *calc_dirtax.do*, *do_household.do*, *do_distanalysis.do*, *indirect_incidence.do*, *labor.do* and *quaid.do*) in that folder.
- Create a sub-folder in “LATAx” called “Params_Data” and place the parameters modules (*params_2008*, *params_2010* *params_20101*).
- Create sub-folders in “LATAx” called “Input_Data” and “Output_Data” in which the input datasets for use in the program and output datasets created by the program will be stored, respectively.

Once this is done, you should have a folder structure as illustrated in Figure 2⁴:

Figure 2. Recommended LATAx folder structure



Once the folders are created and the program used, the next stage is to put the input datasets in to the “Input_Data” folder. Instructions on how to create the input data can be found in Section 6 of this instruction manual. Alternatively, you can download datasets based on the 2008 version of the ENIGH from the LATAx website (see above).

Last, but not least, load Stata and open up the *main_module.do* and *params_2008.do* parameters file in the do file editor. The program is now ready to be used. Section 4 explains how to use it, using our worked example.

⁴ If you are planning to use LATAx for analysis in multiple countries it is recommended that you create sub-folders in the “LATAx” for each country (the name should contain no spaces: e.g. “ElSalv” for El Salvador). Then place the Input_Data, Output_Data and Params_Data folders (and relevant data and do files) for each country in that country sub-folder.

4 Running the model: interface module and parameters

This section uses a worked example – simulating the 2010 Mexican tax reforms – to explain how to use the LATAX program. It begins with Stata open, and the programs *main_module.do* and *params_2008.do* open in the do file editor (see Section 3).

4.1 Setting up the parameters files

The first step is to get the systems parameters files ready for use. LATAX *always* requires a baseline system (in this practise exercise the pre-reform 2008 system – *params_2008*), and two reforms systems (here, the implemented and initially proposed systems for 2010 – *params_2010* and *params_20101*, respectively), and parameters files need to be created for each of these systems.

The parameters files contain information on the various tax rates, thresholds, exemptions and allowances, required to model the different parts of the tax system. A description of every parameter can be found in appendix B; here we focus on implementing the changes to the parameters required to model the 2010 tax reforms, using the 2008 tax system as a baseline. Open the *params_2008* do file, and immediately save it as *params_20100*. We will use this as the 2010 parameters file (so as not to save over the pre-made *params_2010* file).

The first set of parameters, *OVAT**, *OBVAT** and *ODUTIES** refer to the VAT and duty rates as they were when the survey data being used was collected. The default values are the 2008 rates, so if you are using 2008 data – which is the default presently – no changes need to be made here. If, on the other hand you are using a more up to date dataset (2010, or 2012, which will be released in 2013), you will need to update these values to the correct rates for that year.

The first thing that does need changing is the actual VAT rates, stored in the *VAT** and *BVAT** scalars. Initially, 14 goods categories are either exempt from VAT or subject to a zero-rate so that *VAT** and *BVAT** are equal to 0. The next 16 goods are subject to either a VAT rate of 15% ($VAT=0.15$) or 10% for certain border areas ($BVAT=0.1$). The remaining 30 goods represent informally purchased goods and services for which we assume no VAT is paid.⁵ The 2010 reforms raised the rate of VAT from 15% to 16%, and the special rate of VAT applying in border areas from 10% to 11%. To make this change, replace $VAT*=0.15$ with $VAT*=0.16$ for $* = 15$ to 30; and $BVAT* = 0.1$ with $BVAT*=0.11$ for $*=15$ to 30. Figure 3 shows what the parameters should look like after changes have been made.

The 2010 reforms also saw an increase in certain duty (*IEPS*) rates and the introduction of a 3% tax on telecommunications services. The *DUTIES** scalars store the duty rates, expressed as a percentage of the pre-tax price.⁶ By default, the goods subject to duties are listed as goods 25 – 30. To simulate the 2010 reforms, make the following changes:

- $DUTIES25 = 0.03$ (from 0.00): the 3% telecoms tax
- $DUTIES26 = 0.265$ (from 0.25): the increase in tax on wine and alcohol <20% alcohol
- $DUTIES27 = 0.3$ (from 0.2): the increase in gaming/lottery tax

⁵ In principle,

⁶ In reality, duties in Mexico are a mix of specific duties per unit or by volume, and ad-valorem (percentage) duties. Because of poor quality information in the ENIGH survey on quantities purchased, we used “average” ad valorem rates calculated by CEFPE. It would be possible to amend the tax simulator to allow for ad valorem and specific duty rates by amending the consumption input dataset to include quantities, and the indirect tax parameters and simulation code accordingly.

- DUTIES28 = 0.53 (from 0.5): the increase in taxes on spirits > 20% alcohol
- DUTIES29 = 0.265 (from 0.25): the increase in tax on beer
- DUTIES30 = 1.64 (from 1.60): the increase in tax on cigarettes

Figure 4 shows the duty rates following these changes.

Figure 3. VAT rates following the 2010 tax reforms

```
***** NEW VAT *****
scalar VAT1 = 0
scalar VAT2 = 0
scalar VAT3 = 0
scalar VAT4 = 0
scalar VAT5 = 0
scalar VAT6 = 0
scalar VAT7 = 0
scalar VAT8 = 0
scalar VAT9 = 0
scalar VAT10 = 0
scalar VAT11 = 0
scalar VAT12 = 0
scalar VAT13 = 0
scalar VAT14 = 0
scalar VAT15 = 0.16
scalar VAT16 = 0.16
scalar VAT17 = 0.16
scalar VAT18 = 0.16
scalar VAT19 = 0.16
scalar VAT20 = 0.16
scalar VAT21 = 0.16
scalar VAT22 = 0.16
scalar VAT23 = 0.16
scalar VAT24 = 0.16
scalar VAT25 = 0.16
scalar VAT26 = 0.16
scalar VAT27 = 0.16
scalar VAT28 = 0.16
scalar VAT29 = 0.16
scalar VAT30 = 0.16
```

Figure 4. Duties rates following the 2010 tax reforms

```
scalar DUTIES25 = 0.03
scalar DUTIES26 = 0.265
scalar DUTIES27 = 0.30
scalar DUTIES28 = 0.53
scalar DUTIES29 = 0.265
scalar DUTIES30 = 1.64
```

The 2010 tax reforms also involved changes to income tax rates. The tax rates are only one of the income tax parameters that can be easily varied in LATAX: users can also easily change the thresholds, exemptions, the number of rates, and tax credit amounts, information on which is available in appendix B. To change the tax rates (RATE6, RATE7 and RATE8) as shown in Figure 5. This is the final change to model for the 2010 tax system so save the params_20100 do file.

Figure 5. Income tax parameters following 2010 tax reforms

```
***** INCOME TAX THRESHOLDS *****  
  
scalar NUMBANDS = 8  
  
scalar BAND0 = 0  
scalar BAND1 = 5952.84  
scalar BAND2 = 50524.92  
scalar BAND3 = 88793.04  
scalar BAND4 = 103218.00  
scalar BAND5 = 123580.20  
scalar BAND6 = 249243.48  
scalar BAND7 = 392841.96  
scalar BAND8 = 1000000000  
  
***** INCOME TAX RATES *****  
  
scalar RATE1 = 0.0192  
scalar RATE2 = 0.064  
scalar RATE3 = 0.1088  
scalar RATE4 = 0.16  
scalar RATE5 = 0.1792  
scalar RATE6 = 0.2136  
scalar RATE7 = 0.2352  
scalar RATE8 = 0.3
```

The changes to the tax system parameters to simulate the originally proposed 2010 system can be made in the same manner (the full list of reforms is available in Appendix A). But, to save time for this practise exercise you can just use the pre-defined `params_20101` system.

4.2 Setting up main_module options

The next step is to set the instructions in `main_module.do` to ensure the correct input files are used, output files are named and stored correctly, and the right program options are used. This can be done either by making changes to `main_module.do` or using interactive commands typed into the Stata command line. We first describe how to set the options by amending `main_module.do` directly – which is particularly useful when changing the default options –, and then how to set things up using the interactive commands.

Amending main_module.do directly

Figure 6 shows the default directory and filename options in LATAx when downloaded from the internet.

The first thing you might want to do is change the location of the LATAx program and data files. To do this, you amend the `taxsim`, `in`, `out` and `param` globals. For instance, because LATAx has been designed to be a multiple country model, the existing default input, output and parameters directories include space for a country sub-directory:

e.g. global in "C:\LATAx\\"\$country"\Input_Data"

If you are planning to use the model only for Mexico and have created the folders as set out in Section 3, simply delete `\$country\` from each of the `in`, `out` and `param` globals.

Figure 6. Default directory and filename options

```
/** MAIN PROGRAM **/  
program set_defaults  
  
* Instructions  
  
* 1. assign the root/program directory.  
* 2. assign the input, output, and parameter directories.  
* 3. assign the name for the logfiles.  
* 4. assign the name of the input dataset.  
* 5. assign the name of the output dataset.  
  
* Assign default directories  
  
global taxsim      "C:\LATAx\Do_Files"  
global in          "C:\LATAx\%country%\Input_Data"  
global out         "C:\LATAx\%country%\Output_Data"  
global param       "C:\LATAx\%country%\Params_Data"  
  
* Assign default log files  
global logfiledecile "decile.txt"  
global logfiledecile1 "decileburden.txt"  
global logfilehhtype "hhtype.txt"  
global logfilerevenue "revenue.txt"  
  
* Assign default input and output files  
  
global indata_hh "Input_Data_hh"  
global indata_ad "Input_Data_ad"  
global indata_co "Input_Data_co"  
  
global outdata_ad "Output_adlevel"  
global outdata_hh "Output_hhlevel"
```

Second, you might also want to change the name of the various Stata logfiles created by LATAx.

logfiledecile	is the logfile used to record changes in tax payments / net incomes by household income and expenditure decile groups
logfiledecile1	is the logfile used to record the proportion of the gains/losses following a reform accruing to each household income and expenditure decile group
logfilehhtype	is the logfile used to record changes in tax payments / net incomes by household type (as defined in the input data)
logfilerevenue	is the logfile used to record the changes in tax revenues following a reform

In this example, as this is a practise, let us suppose you just add `_practise` to the end of every filename (but before the `.txt`).

Third, the globals **input_hh**, **input_ad** and **input_co** define the names of the input files that contain basic household demographics, adult demographics and incomes, and expenditure information, respectively. These should be amended if you have named the input files differently, or if you want to use a different version of an input file (for instance, one that corrects for under-recording of income or expenditure differently than the standard file). In this example, let's leave the filenames unchanged.

Fourth, the globals **outdata_ad** and **outdata_hh** define the names of the output files that store adult and household level measures of the impact of the reforms. Again, let us rename these by adding **_practise** at the end. Figure 7 shows what you now should have.

Figure 7. Amended directory and filename options

```
/** MAIN PROGRAM **/  
program set_defaults  
  
* Instructions  
  
* 1. assign the root/program directory.  
* 2. assign the input, output, and parameter directories.  
* 3. assign the name for the logfiles.  
* 4. assign the name of the input dataset.  
* 5. assign the name of the output dataset.  
  
* Assign default directories  
  
global taxsim    "C:\LATA\Do_Files"  
global in        "C:\LATA\Input_Data"  
global out       "C:\LATA\Output_Data"  
global param     "C:\LATA\Params_Data"  
  
* Assign default log files  
global logfiledecile "decile_practise.txt"  
global logfiledecile1 "decileburden_practise.txt"  
global logfilehhtype "hhtype_practise.txt"  
global logfilerevenue "revenue_practise.txt"  
  
* Assign default input and output files  
  
global indata_hh "Input_Data_hh"  
global indata_ad "Input_Data_ad"  
global indata_co "Input_Data_co"  
  
global outdata_ad "Output_adlevel_practise"  
global outdata_hh "Output_hhlevel_practise"
```

The next stage is to change the various options which tell the program what type of analysis you want to perform (see Figure 8).

The global **disttype** tells LATAx whether to conduct the distributional analysis using a range of equivalence scales to adjust for household size (**disttype=1**) or only the standard equivalence scale (**disttype=2**), where second and subsequent adults or children aged 12 or over are weighted as 0.8 of the first adult, and children aged 11 or under are weighted as 0.5 of an adult. For this practise, set it equal to 1.

The next global, **netincometype**, tells LATAx whether it needs to calculate household net income (**netincometype=1**) or whether it is provided in the household level input dataset **indata_hh** (**netincometype=2**). If you are using either Mexican input data downloaded from the LATAx website or created input data from ENIGH using the do-files provided (see Section 6), leave the this global at its default value (=2). If you have created your own input dataset from data that does not include net income measures, change to =1.⁷

identlist lists the variables used to identify households. The default is “**folioviv foliohog**” for Mexico.⁸ Only change this if you have created your own input dataset with different household identifiers.

PTO

⁷ Also set to 1 if using LATAx with data for El Salvador to model reforms in that country.

⁸ Change this to “hhid” if using the model for analysis in El Salvador using El Salvadoran data.

Figure 8. Amended control variables, labor supply and consumer demand options

```
/** CONTROL VARIABLES **/

* 1 = full range of distributional analyses, 2 = reduced range
global disttype = 1

* 1 = use calculated net income, 2 = use net income from input
global netincometype = 2

* Set the Household Identifier List

global identlist "folioviv foliohog"

/** MODULES INCLUDING BEHAVIORAL RESPONSE MARGINS **/

/** 1) LABOR SUPPLY MODEL **/

* do you want to run labor model (1 = Yes, 2 = No)
* DEFAULT SET TO 2 (NO)

global labmod = 1

* Others

global logfilelabor "labor_practise.txt"

/** 2) CONSUMPTION MODEL **/

* do you want to run consumption model? (1 = Yes, 2 = No)
* DEFAULT SET TO 2 (NO)

global consmod = 1

* do you want to reduce spending in response to changes in dire

global inc_aff_spend = 1

* Others

global indata_demand "demand_params.dta"
global indata_prices "demand_prices.dta"

global logfiledemand_welfare "decile_demand_practise.txt"
global logfiledemand_spending "spending_demand_practise.txt"
global logfiledemand_revenue "revenue_demand_practise.txt"
```

Next, are options which tell LATAx whether to perform behavioural analysis.

The global **labmod** determines whether labor supply modelling (using assumed participation and 'hours'/earnings elasticities) is carried out. The default is for it not to be (=2). Amend this to =1 so that such modelling is done. **logfilelabor** is the name of the file that records the outputted labor supply effects. For this practise, change this to "**labour_practise.txt**".

The global **consmod** determines whether consumer demand modelling is carried out. The default is for it not to be (=2). Amend this to =1 so that such modelling is done. **inc_aff_spend** tells LATAx whether or not to assume that changes in net income following changes in income tax or social security contributions feed into changes in consumption. The default is that they do not (=2); changing this global to = 1 means that they do (the percentage change in total expenditure is assumed to be the same as the percentage change in net income).

The next set of globals determines input files for the consumer demand modelling (**indata_demand** contains the demand system parameters, and **indata_prices** contains data on the prices of different goods) and the logfiles used for outputs (beginning with **logfiledemand_**). In this practise exercise, add **_practise** to the end of each logfile file name. If using the default demand system and input data included with LATAx for Mexico, *do not* change the input datasets, or the remaining consumer demand globals (**NUMGOODSQUAIDS** to **catglist60**). Appendix D describes what to do if you wish to estimate a new demand system and incorporate that into the model.

The final behavioural model is that which allows firms to pass less than 100% of changes in indirect taxes (IVA and IEPS) through to consumer prices (see Figure 9). The default is that this option is turned off (**indir=2**), but in for this practise exercise, let us turn it on (**indir=1**).

Figure 9. Amended control variables, labor supply and consumer demand options

```

/**** 3) INDIRECT TAX INCIDENCE ****/

* do you want to estimate distributional effects with alternat
* DEFAULT SET TO 2 (NO)

global indir = 1

* proportion of pass-through
global prop_prices = 0.75

* proportion not passed through borne by wages
global prop_wages = 0.75

* output data for the different pass-through of indirect tax
global outdata_ad_ind "out_adlevel_ind_p75_w75_practise"
global logfiledecile_ind "decile_ind_p75_w75_practise.txt"
global logfilehhtype_ind "hhtype_ind_p75_w75_practise.txt"
global logfilerevenue_ind "revenue_ind_p75_w75_practise.txt"
|
* control parameter (don't change)
global indic_incidence = 1

```

If this option is turned on, you must tell LATAx the proportion of any indirect tax changes that are passed through to prices (**prop_prices=**), and the proportion of the indirect tax change that is borne by firms (i.e. not passed through to prices) that is ultimately borne by workers in the form of lower wages as opposed to being borne by shareholders (**prop_wages=**). For this practise exercise, let us assume that both proportions are 0.75. In other words, 75% of any changes in indirect tax will be passed through to prices; of the 25% borne by firms, 75% will be borne by workers in the form of lower wages.

The next globals define the adult-level output file and the log files for the analysis with less-than-full pass through of indirect tax (**outdata_ad_ind**, **logfiledecile_ind**, **logfilehhtype_ind**

and logfilerevenue_ind). It is a good convention to ensure that the names include an “_ind” element to make sure you don’t overwrite the files created for the analysis assuming full pass-through. For this practise lets also add “_p75_w75_practise” to the end of each file name to indicate that 75% of indirect tax changes are passed into prices, 75% of the tax borne by firms is passed into wages, and that this is a practise exercise.

Never change the final “control parameter” (**indic_incidence=1**).

The final step is to run the program. To do this:

- Click the “execute (do)” button on the right hand side of the do file editor toolbar. This runs the code.
- Then, in the main Stata command line type (without quotations) “set_systems” followed by the numbers of the baseline and two reform systems. In this instance “set_systems 2008 20100 20101”
- Then, type “set_defaults”
- Then type “simulator”

The program is now running. Once it has finished you can look at outputs (see Section 5).

Programming LATAX using Stata command line programs

If you decide to set up LATAX using Stata command line commands rather than by amending the *main_module.do* file, the first stage is to click the “execute (do)” button on the right hand side of the do file editor toolbar (ensuring the *main_module.do* file is the active do file displayed on screen before doing this). Then, in the main Stata command line type (without quotations) “set_systems” followed by the numbers of the baseline and two reform systems. In this instance “set_systems 2008 2010 20101”, followed by “set_defaults”. You are now ready to start telling LATAX directory locations, filenames and selecting program options. A full list of the optional command-line commands can be found in appendix B. Here we describe those needed to set the program up for the practise exercise, as described above.

The first thing to do is to set the directories for the input, output and parameters data. You do this by typing the following commands into the Stata command prompt and pressing enter after each line:

set_inputdir “C:/LATAX/Input_Data”

set_outputdir “C:/LATAX/Output_Data”

set_paramdir “C:/LATAX/Params_Data”

This overwrites the default options only until the defaults are re-set, so is handy if you want to make a temporary change to an output folder.

Given that we are using the default input files, there is no need to change these. But we do want to change the output file names. Input and execute the following commands:

set_adout “output_adlevel_practise”

set_hhout “output_hhlevel_practise”

and the name of the log files for distributional and revenue analysis:

set_logdec “decile_practise.txt”

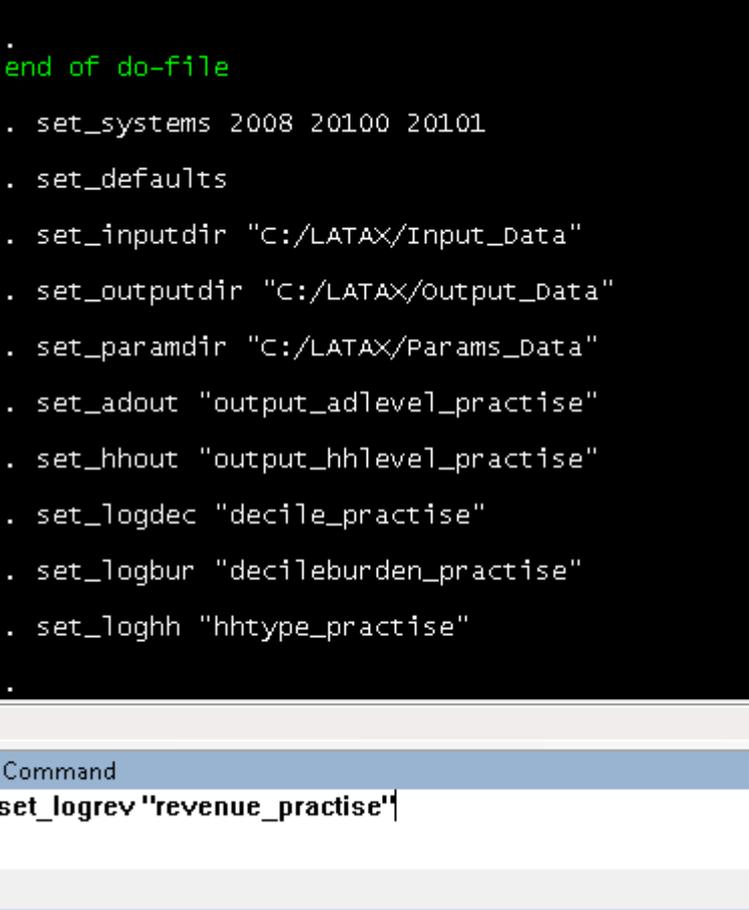
set_logbur “decileburden_practise.txt”

set_loghh "hhtype_practise.txt"

set_logrev "revenue_practise.txt"

Figure 10 shows how the (bottom of) the main Stata window and command prompt line should look when you are entering the last of these commands.

Figure 10. Directories and filenames inputted interactively using Stata commands



```
.  
end of do-file  
. set_systems 2008 20100 20101  
. set_defaults  
. set_inputdir "C:/LATAX/Input_Data"  
. set_outputdir "C:/LATAX/Output_Data"  
. set_paramdir "C:/LATAX/Params_Data"  
. set_adout "output_adlevel_practise"  
. set_hhout "output_hhlevel_practise"  
. set_logdec "decile_practise"  
. set_logbur "decileburden_practise"  
. set_loghh "hhtype_practise"  
. .  
Command  
set_logrev "revenue_practise"
```

Next, the options that tell the program which type of analysis to conduct should be set. First, we set which type of distributional analysis to conduct, and tell the program to use net income from the input files (rather than calculating it).

set_disttype 1

set_netinctype 2

Before turning on labour supply modelling and setting the location of the output file:

set_labmod 1

set_labour "labor_practise.txt"

And then turning on consumer demand modelling and setting the location of the input and output files required for this modelling.

set_consmod 1

set_changewithinc 1

set_demandparam "demand_params.dta"

`set_demandprice "demand_prices.dta"`

`set_demandwelfare "decile_demand_practise.txt"`

`set_demandspending "spending_demand_practise.txt"`

`set_demandrevenue "revenue_demand_practise.txt"`

Figure 11 shows the inputting of the options for these options.

Figure 11. Setting labor supply and consumption options using Stata commands

```
. set_disttype 1
. set_netinctype 1
. set_labmod 1
. set_labout "labor_practise.txt"
. set_consmo 1
. set_changewithinc 1
. set_demandparam "demand_params.dta"
. set_demandprice "demand_prices.dta"
. set_demandwelfare "decile_demand_practise.txt"
. set_demandspending "spending_demand_practise.txt"
.
```

Command
`set_demandrevenue "revenue_demand_practise.txt"`

The last set of options to program up are those relating to less-than-full pass through of indirect tax changes in to prices. These tell the program whether to run the module with less than full pass through (`set_indir 1` means yes), the proportion of the tax change passed through to prices (`set_propprices`) and the proportion of that not passed through that impacts on wages (`set_propwages`) as well as the output files and log files. Figure 12 shows these commands once they have been entered.

Once this is done, all necessary commands have been input and the model can be run by typing "simulator" into the Stata command line and pressing enter.

Section 5 explains how to interpret the results of the simulator once they have been outputted. Before that, Section 4.3 provides information for readers who are interested in changing the built-in elasticities of labor supply used in the labor supply model built into LATAX: other readers can skip to Section 5.

Figure 12. Setting less-than-full pass through indirect tax options interactively using Stata commands

```
. set_indir 1
. set_propprices 0.75
. set_propwages 0.75
. set_adout_ind "out_adlevel_ind_p75_w75_practise.dta"
. set_logdec_ind "decile_ind_p75_w75_practise.txt"
. set_loghh_ind "hhtype_ind_p75_w75_practise.txt"
. set_logrev_ind "revenue_ind_p75_w75_practise.txt"
.
```

Command
simulator

4.3 Changing labour supply elasticities

The labor supply model built in to LATAX allows for elasticities of employment and hours of work / earnings to vary by sub-group of the population. To prevent the assumptions for different groups cluttering up the main_program interface module, these have been hard-coded into labor.do. Users who wish to change these (for instance, who want to assume labor supply is more or less responsive than the assumptions built into LATAX as standard, which are broadly on the high-side for developed countries but were felt to be a reasonable central guess where responsiveness could be greater due to the ability to work informally) will therefore need to open labor.do do file and make changes directly there. This is easy to do if one wants to allow elasticities to vary by basic features such as sex, number of children, age, position in the earnings distribution etc, as these variables are already defined in the standard LATAX input datasets or calculated in the labor.do program.

The labor supply module uses a set of elasticities for the extensive (participation/employment) decision (**partelast**) and the intensive (hours) decision (**hourselast**). A higher elasticity implies more behavioral response to a change in tax rates, and, for instance, a lower yield from tax increases (or a smaller cost for tax cuts). To edit the elasticities, the following blocks of code need to be amended:

```
gen hourselast = 0.05
replace hourselast = 0.1 if sex==2 & menores>1
replace hourselast = 0.1 if percgross>90
replace hourselast = 0.2 if percgross==100
```

```

gen partelast = 0.05
replace partelast = partelast*2 if percgross>40 & percgross<=60
replace partelast = partelast*3 if percgross>20 & percgross<=40
replace partelast = partelast*4 if percgross>0 & percgross<=20
replace partelast = partelast*2 if menores>1 & sex==2

```

5 Interpreting outputs of model: distribution, revenue, behavior

Once the program has finished running (you will see “end of do file” at the bottom of the Stata output window), you can then start looking at the outputs of the model. These outputs will consist of a number of Stata .dta files and text log files. When looking for summary distributional, revenue and behavioral measures of reforms, the log files provide the necessary information, therefore they are the focus in this section. Information on the variables contained in the output files can be found in Appendix C.

5.1 Main log files from the ‘no behavioural response’ simulations

The outputs of the simulation will be stored according to the directory (**out**) and filenames chosen when setting up the LATAX simulation. Four log files are produced for the no-behavioural-response simulations.

Then first records the revenue effects of the reform, and is stored in the filename declared in **logfile_revenue** (in this case revenue_practise). Figure 13 shows the format of this output.

Figure 13. The revenue log file

```

name: <unnamed>
log: C:\LATAX\output_data\revenue_practise.txt
log type: text
opened on: 18 Dec 2012, 20:44:43

Revenue raised from the tax reforms

Tax | Reform System 20100 | Reform System 20101
-----|-----|-----
ISR | $ 5,987,545,600 | $ 8,471,142,912
Soc. Sec. | $ 0 | $ 0
IVA | $ 10,888,586,240 | $ 38,020,403,200
IEPS | $ 3,064,007,424 | $ 4,083,248,640
Total | $ 19,940,139,008 | $ 50,574,794,752
name: <unnamed>
log: C:\LATAX\output_data\revenue_practise.txt
log type: text
closed on: 18 Dec 2012, 20:44:43

```

The log reports the change in revenue from each tax in *Mexican Pesos* following the two sets of reforms analysed. A positive value means an increase in revenue and a negative value means a decrease in revenue following the reforms.

The analysis of the impact of the reforms across the income and expenditure distributions is recorded in the log file declared in **logfile_decile** (in this case decile_practise). Gains to

households are shown as positive values and losses are shown as negative values. This log file includes multiple tables that show the distributional impact of the reform in slightly different ways:

- Using expenditure or income as the variable by which to rank households into decile groups from poorest to richest
- Using expenditure or income to divide cash gains/losses into proportional gains/losses
- Using measures of total income and expenditure that include non-monetary amounts such as payment-in-kind and imputed rent, and using measures of monetary income and expenditure only
- And using different equivalence scales to adjust incomes and expenditures for household size and composition when ranking households into decile groups.

Thus for instance, the first two tables show the impact of the first set of reforms (in this case system 20100) in cash terms (the first table in Figure 14) and as a proportion of total expenditure (**texp**) (the second table in Figure 14), for deciles of the total expenditure (**texp**) distribution, when households are ranked using an equivalence scale whereby expenditure (and income) is divided by the number of household members (**es11**). This is what the heading “IMPACT OF REFORMS AS A PROPORTION OF **texp** USING DECILES OF **texp** USING EQUIVALENCE SCALE **es11** FOR REFORM SYSTEM 20100” means.

Figure 14. Interpreting the decile log file

name: <unnamed>
 log: C:\LATA\output_Data\decile_practise.txt
 log type: text
 opened on: 19 Dec 2012, 11:39:18

IMPACT OF REFORMS IN CASH TERMS USING DECILES OF **texp** USING EQUIVALENCE SCALE **es11** FOR REFORM SYSTEM 20100

Decile	texp	ISR	Soc. Sec.	IVA	IEPS	TOTAL
Poorest	\$ 38,626	\$ -1	\$ 0	\$ -86	\$ -14	\$ -101
Decile 2	\$ 57,940	\$ -8	\$ 0	\$ -145	\$ -30	\$ -183
Decile 3	\$ 70,275	\$ -10	\$ 0	\$ -190	\$ -49	\$ -249
Decile 4	\$ 78,243	\$ -17	\$ 0	\$ -220	\$ -63	\$ -300
Decile 5	\$ 90,825	\$ -34	\$ 0	\$ -273	\$ -86	\$ -394
Decile 6	\$ 102,458	\$ -67	\$ 0	\$ -323	\$ -107	\$ -496
Decile 7	\$ 116,323	\$ -121	\$ 0	\$ -387	\$ -124	\$ -632
Decile 8	\$ 135,926	\$ -209	\$ 0	\$ -478	\$ -154	\$ -842
Decile 9	\$ 166,351	\$ -438	\$ 0	\$ -637	\$ -197	\$ -1,272
Richest	\$ 324,373	\$ -1,340	\$ 0	\$ -1,339	\$ -324	\$ -3,003
Average	\$ 118,121	\$ -224	\$ 0	\$ -408	\$ -115	\$ -747

IMPACT OF REFORMS AS A PROPORTION OF **texp** USING DECILES OF **texp** AND EQUIVALENCE SCALE **es11** FOR REFORM SYSTEM 20100

Decile	texp	ISR	Soc. Sec.	IVA	IEPS	TOTAL
Poorest	\$ 38,626	-0.00%	0.00%	-0.22%	-0.04%	-0.26%
Decile 2	\$ 57,940	-0.01%	0.00%	-0.25%	-0.05%	-0.32%
Decile 3	\$ 70,275	-0.01%	0.00%	-0.27%	-0.07%	-0.35%
Decile 4	\$ 78,243	-0.02%	0.00%	-0.28%	-0.08%	-0.38%
Decile 5	\$ 90,825	-0.04%	0.00%	-0.30%	-0.09%	-0.43%
Decile 6	\$ 102,458	-0.07%	0.00%	-0.31%	-0.10%	-0.48%
Decile 7	\$ 116,323	-0.10%	0.00%	-0.33%	-0.11%	-0.54%
Decile 8	\$ 135,926	-0.15%	0.00%	-0.35%	-0.11%	-0.62%
Decile 9	\$ 166,351	-0.26%	0.00%	-0.38%	-0.12%	-0.76%
Richest	\$ 324,373	-0.41%	0.00%	-0.41%	-0.10%	-0.93%
Average	\$ 118,121	-0.11%	0.00%	-0.31%	-0.09%	-0.51%

The first column shows the (*unequalised*) average expenditure (or income) for each decile group, the next four columns show the average (mean) gain or loss for households in each decile group

from changes in the amount of income tax, social security, VAT/IVA and duties/IEPS paid. The last column shows the average total gain or loss for households in each decile group.

Scrolling down, the next two tables show the impact of the second set of reforms (system 20101) on the same basis (i.e. % of total expenditure, total expenditure decile groups and a per-capita equivalence scale).

The impact of the two sets of reforms is then shown as a proportion of monetary expenditure (**mexp**), total income (**tinc**) and monetary income (**minc**), in turn, again using total expenditure decile groups and a per-capita equivalence scale.

Next, outputs are produced showing gains/losses as a proportion of total expenditure (**texp**), monetary expenditure (**mexp**), total income (**tinc**) and monetary income (**minc**), in turn, this time using monetary expenditure decile groups and per capita equivalence scales. This is then repeated for total income decile groups and monetary income decile groups.

Lastly, everything is repeated for two other equivalence scales. First, an equivalence scale where the first adult is given a weight of 1, subsequent adults and children aged 12 or over are given a weight of 0.8, and children aged 11 or under are given a weight of 0.5 (**es85**). Then, an equivalence scale where the first adult is given a weight of 1, subsequent adults and children aged 12 or over are given a weight of 0.5, and children aged 11 or under are given a weight of 0.3 (**es53**). This latter scale is that used in the UK and is termed the “modified OECD scale”. Because a larger proportion of expenditure in Mexico is likely to be on food, and other goods that cannot be easily be shared, the modified OECD scale may result in too-small adjustments to the income for larger households when ranking them into decile groups. The **es85** scale allows for some (but fewer) economies of scale, whilst the **es11** scale implies no economies of scale (and that children ‘cost’ the same as adults).

Figure 15 shows the impact of the reform system 20100 as a percentage of total income across the total income distribution.

Figure 15. Dividing gains/losses by income or expenditure?

IMPACT OF REFORMS IN CASH TERMS USING DECILES OF tinc USING EQUIVALENCE SCALE es11 FOR REFORM SYSTEM 20100

Decile	tinc	ISR	Soc. Sec.	IVA	IEPS	TOTAL
Poorest	\$ 33,795	\$ 0	\$ 0	\$ -111	\$ -18	\$ -129
Decile 2	\$ 56,039	\$ -0	\$ 0	\$ -159	\$ -33	\$ -193
Decile 3	\$ 71,422	\$ -0	\$ 0	\$ -204	\$ -52	\$ -256
Decile 4	\$ 87,847	\$ -1	\$ 0	\$ -226	\$ -63	\$ -290
Decile 5	\$ 106,375	\$ -5	\$ 0	\$ -280	\$ -84	\$ -369
Decile 6	\$ 124,121	\$ -26	\$ 0	\$ -332	\$ -103	\$ -461
Decile 7	\$ 140,826	\$ -49	\$ 0	\$ -374	\$ -119	\$ -542
Decile 8	\$ 174,744	\$ -174	\$ 0	\$ -483	\$ -164	\$ -821
Decile 9	\$ 226,157	\$ -391	\$ 0	\$ -666	\$ -205	\$ -1,262
Richest	\$ 475,085	\$ -1,597	\$ 0	\$ -1,245	\$ -307	\$ -3,149
Average	\$ 149,634	\$ -224	\$ 0	\$ -408	\$ -115	\$ -747

IMPACT OF REFORMS AS A PROPORTION OF tinc USING DECILES OF tinc AND EQUIVALENCE SCALE es11 FOR REFORM SYSTEM 20100

Decile	tinc	ISR	Soc. Sec.	IVA	IEPS	TOTAL
Poorest	\$ 33,795	0.00%	0.00%	-0.33%	-0.05%	-0.38%
Decile 2	\$ 56,039	-0.00%	0.00%	-0.28%	-0.06%	-0.34%
Decile 3	\$ 71,422	-0.00%	0.00%	-0.29%	-0.07%	-0.36%
Decile 4	\$ 87,847	-0.00%	0.00%	-0.26%	-0.07%	-0.33%
Decile 5	\$ 106,375	-0.00%	0.00%	-0.26%	-0.08%	-0.35%
Decile 6	\$ 124,121	-0.02%	0.00%	-0.27%	-0.08%	-0.37%
Decile 7	\$ 140,826	-0.03%	0.00%	-0.27%	-0.08%	-0.39%
Decile 8	\$ 174,744	-0.10%	0.00%	-0.28%	-0.09%	-0.47%
Decile 9	\$ 226,157	-0.17%	0.00%	-0.29%	-0.09%	-0.56%
Richest	\$ 475,085	-0.34%	0.00%	-0.26%	-0.06%	-0.66%
Average	\$ 149,634	-0.07%	0.00%	-0.28%	-0.08%	-0.42%

Comparing the outputs in Figures 14 and 15, one can see how the distributional impact of reforms can look significantly different when households are ranked into decile groups differently, and especially, when gains/losses are expressed as a proportion of income rather than expenditure. In this instance, the reforms (system 20100) look to be significantly more progressive (i.e. the rich bear more of the burden) when using total expenditure decile groups and dividing losses by total expenditure (Figure 14) than when using total income decile groups and dividing losses by total income (Figure 15). In particular, the changes to IVA (VAT) appear to be progressive when expressed as a fraction of expenditure over the expenditure distribution (losses equivalent to 0.41% for the ‘richest’ 10% and 0.22% for the ‘poorest’ 10%, on average), but slightly regressive when expressed as a fraction of income over the income distribution (losses equivalent to 0.26% for the ‘richest’ 10% and 0.33% for the ‘poorest’ 10%, on average). What figures give a more meaningful picture of the distributional impact of the reforms?

It turns out that the issue of how to group households into decile groups is a separate issue from whether to divide gains/losses by income or expenditure. A case can be made to group households into decile groups based on their income or their expenditure (both are volatile and it is not clear what better captures households’ underlying purchasing-power), but, as explained in box 5.1, it is most meaningful to divide gains/losses from changes to direct taxes (ISR and Social Security taxes) by *income* and changes in indirect taxes (IVA and IEPS) by *expenditure*. When examining reforms that affect both direct and indirect taxes (such as system 20100), one should consider the impact both as a percentage of income and a percentage of expenditure, giving more weight to the former if direct tax changes are most significant, and the latter if indirect tax changes are most significant. However, sometimes it will, necessarily, be difficult to determine whether a reform is progressive or regressive.⁹

Box 5.1 Income or expenditure to assess the distributional impact of tax reforms?

The literature on tax reform analysis has noted that whether income or expenditure is used to rank households and to calculate proportional gains/losses can have a major impact on the assessed distributional impacts (Abramovsky et al (2011), IFS et al (2011)). For instance, the VAT systems operating in most countries look regressive when VAT paid is expressed as a fraction of income over the income distribution, but slightly progressive when VAT paid is expressed as a fraction of expenditure over the income (or expenditure) distribution. This would appear to present significant difficulties to the analyst: results are highly sensitive to how one decides to calculate the proportional burden of a tax. But can economic reasoning guide us to the *right* way to analyse reforms?

First up, it is not simply a question of using income or expenditure. There are two distinct questions which may have different answers:

- How should one rank households as rich or poor?
- How should one assess the proportional change in tax payments resulting from a reform?

⁹ As is explained in Box 5.1, what is really needed to determine reform progressivity is lifetime income and expenditure (or a good proxy for this) which is not available.

The answers to both are strongly linked to the long-term or 'life-time' distributional impact of a tax change, but in different ways.

First, how should households be ranked?

The issue at stake here is whether a household's position in the income distribution or expenditure distribution gives a better indication of whether they are rich or poor. In order to assess this, one must first understand that household surveys generally pick up a 'snapshot' measure of income or expenditure (e.g. income in the last month, or spending on different types of items in periods ranging from one week to one year). But such a short-term measure might not accurately reflect the living standards of the household in either the short or long run. For instance, households with low incomes may be able to use borrowings, savings or previously purchased durable goods to maintain their living standards, at least in the short run.

Many economists have argued that households should be ranked by their consumption as this takes account of such 'smoothing' of income shocks (Porterba, 1989; Meyer and Sullivan, 2003, 2004, 2008, 2011). The argument for using consumption is particularly persuasive if we believe households smooth their consumption over long periods of time and we are concerned with the long-term distributional impact of a policy change.

This might suggest a preference for using expenditure to rank households. But expenditure is not the same as consumption: expenditure captures the purchase costs of durable goods like cars, whereas consumption captures the flow of benefits from these goods. Like income, expenditure may be volatile, with households purchasing certain items infrequently, especially larger durable goods such as motor vehicles or new kitchens (but also food if they bulk-purchase). Excluding durable goods from the measure of expenditure removes much of this problem but introduces a new one: you may rank households incorrectly if they devote different proportions of their budgets to durable goods.

It is therefore not clear whether expenditure represents a better measure of a household's living standards than income: both are volatile, and furthermore, both suffer significant measurement error in surveys. For this reason it is worthwhile conducting analysis ranking households both according to their position in the income distribution and in the expenditure distribution.

Second, how should we assess proportional gains or losses?

The best way to understand why gains or losses should be expressed as a fraction of income for direct taxes and as a fraction of expenditure for indirect taxes is to use some hypothetical tax changes. As this issue most often comes up when analysing changes to indirect taxes, we use examples for VAT that demonstrate how presenting results as a fraction of expenditure is most appropriate (but analogous examples can be used to show how income should be used for direct changes).

Consider the case of a uniform VAT on all goods and services. Over a lifetime, if lifetime income and lifetime expenditure are equal, this can be clearly seen as distributionally neutral^a: as it is imposed on all goods and services at the same rate, it has the same proportional effect on the purchasing power of rich and poor households. VAT payments under such a system would be the same fraction of both lifetime income and lifetime expenditure for rich and poor households.

But suppose, as in reality, we only have information on current income and spending. If VAT payments are presented as a fraction of current expenditure, this distributionally neutral pattern of payments would be found. However, because households with low current income tend to spend more than their income, and those with high current income tend to spend less, showing payments as a fraction of net income will make the uniform VAT look regressive if households are defined as rich or poor based on their current income. On the other hand, if households are defined as rich or poor based on their current expenditure, because households with the lowest spending tend to report incomes that are higher than their spending, and those with high spending tend to report incomes that are lower than their spending, showing VAT payments as a fraction of net income will make the uniform VAT look progressive. That is, a distributionally neutral uniform VAT can be misleadingly labelled progressive or regressive if VAT payments are expressed as a proportion of net income. For this reason, analysis showing VAT payments as a proportion of household expenditure should be considered more informative and will be the focus of discussion in the remainder of this chapter (although results will be presented as a proportion of income in Annex I for the purpose of completeness).

The argument that showing VAT payments as a fraction of income may give a misleading impression of the lifetime distributional impact of VAT is driven by the potential for households to borrow and save, but it does not rely on households being able to borrow freely or have large amounts of savings to draw-down. Neither does it rely on consumers being rational and forward-looking or engaging in optimal consumption smoothing.

To see this, consider a poor household with a long-run income of 100 euros per week but who is currently spending 200 euros per week, funded by drawing down the last of their savings. Furthermore, suppose that the rate of VAT is 25% on all goods and services. The household would pay 40 euros per week in VAT, equal to 20% of their current spending but 40% of their current income. The question is, which measure is a better reflection of the impact of VAT on the household? It is true that their current income is a better measure of their long-run purchasing power than their current expenditure is. But it does not follow that expressing VAT payments as a proportion of current income gives a better measure of the impact of VAT on that long-run purchasing power. This is because when the household is forced to cut their spending back to the level of their long-run income (100 euros per week), the amount of VAT they would pay falls to 20 euros per week. This is equal to 20% of their current and long-run income, and their long-run expenditure of 100 euros per week.

Measuring VAT payments as a percentage of current spending thus gives a more meaningful measure of VAT's distributional impact than measuring payments as a percentage of current income, even in cases where current income is the better measure of long-run living standards. This demonstrates a key point: the best measure of lifetime living standards might be current income for some households, and current spending for other households. This is a separate issue from the arguments in favour of expressing VAT as a percentage of expenditure instead of income, and expressing direct taxes (like Income Tax) as a percentage of income rather than expenditure.

- a. The assumption that lifetime income and expenditure are equal means that we abstract from gifts and bequests. This is for ease of exposition only: the argument with bequests is more complicated but conclusions are unchanged. For example, when assessing the proportional impact of VAT on households that are recipients of gifts and bequests, it seems clear that we would

want to take into account those gifts and bequests when measuring their lifetime resources. We would not, for instance, wish to say that a household with zero income but large expenditures funded by gifts and bequests is hit infinitely hard by VAT. Including bequests and gifts in the lifetime resources of the recipient makes subtracting them from the resources of the giver attractive to avoid the double counting of gifts and bequests. Adding and subtracting gifts and bequests when calculating lifetime resources in this manner means a uniform VAT would be found to be a constant fraction of both lifetime resources (income) and lifetime expenditure, i.e. it would be distributionally neutral as in the case with no gifts and bequests

In paying careful attention to the proportional impact of a set of reforms, one must not forget to pay heed to how the reforms affect households in *cash* as well. For instance, a reduced rate of VAT for food might be progressive in the sense that it represents a larger gain as a fraction of spending for poorer households than richer ones, on average. This is because food represents a larger fraction of the budgets of poorer households. But richer households spend more on food in absolute cash terms, and hence would gain the most from such a policy in cash terms, suggesting it is not a particularly efficient way to redistribute. Thus a more complete picture of the reforms can be obtained by considering both the cash and proportional tables included in the **logfiledecile** log file. Additional information on the percentage of overall gains or losses from reforms that are attributable to each decile group can be found in the log file declared in **logfiledecile1** (in this case `decileburden_practise`), as shown in Figure 16.

Figure 16. Showing how the burden is borne across the distribution

```

name: <unnamed>
log: C:\LATA\output_data\decileburden_practise.txt
log type: text
opened on: 19 Dec 2012, 11:41:13

PERCENTAGE OF ADDITIONAL REVENUE RAISED BY DECILES OF texp USING EQUIVALENCE SCALE es11 FOR REFORM SYSTEM 2010Q

```

Decile	texp	ISR	Soc. Sec.	IVA	IEPS	TOTAL
Poorest	3.27%	0.04%	0.00%	2.12%	1.23%	1.36%
Decile 2	4.91%	0.34%	0.00%	3.55%	2.65%	2.45%
Decile 3	5.95%	0.43%	0.00%	4.66%	4.23%	3.33%
Decile 4	6.63%	0.74%	0.00%	5.40%	5.49%	4.02%
Decile 5	7.68%	1.53%	0.00%	6.70%	7.46%	5.26%
Decile 6	8.67%	2.98%	0.00%	7.91%	9.31%	6.64%
Decile 7	9.85%	5.38%	0.00%	9.49%	10.80%	8.46%
Decile 8	11.51%	9.33%	0.00%	11.73%	13.43%	11.27%
Decile 9	14.09%	19.53%	0.00%	15.63%	17.18%	17.04%
Richest	27.44%	59.70%	0.00%	32.81%	28.22%	40.18%
Average	10.00%	10.00%	0.00%	10.00%	10.00%	10.00%

The first column shows the fraction of total expenditure (in this instance) accounted for by each total expenditure decile group. For instance, the poorest 10% of households on this measure account for 3.27% of spending, and the richest for 27.44%. The next 4 columns show the fraction of gains or losses accruing to each decile group for changes in each of the 4 taxes. In this instance, around 0.04% of the extra income tax revenue is estimated to be raised from the poorest 10% of households, and 59.7% from the richest 10%. Note that positive numbers in this table mean that households in a decile group are sharing (at least a bit) in the overall burden of a tax raise or gaining (at least a bit) from a tax cut; a negative number means that whilst households on average are losing (gaining), households in that decile group are gaining (losing). The last column shows the fraction of gains or losses accruing to each decile group for the tax reforms as a whole.

The last log file for the “no behavioural response” analysis is that declared in **logfilehhtype** (in this case `hhtype_practise`), as shown in Figure 16. This shows the impact of reforms in cash and proportional (i.e. as a percentage of income and expenditure) terms for different types of households as defined in LATAX. The format is effectively the same as for **logfiledecile**. The household types are:

Couple	- Household containing one couple with no child aged under 18
Single	- Household containing a single adult with no child aged under 18
Couple, children <18	- Household containing one couple and one or more child aged under 18
Single, children <18	- Household containing a single adult and one or more child aged under 18
Pensioner, single	- Household containing a single person aged over the pension age
Pensioner, couple	- Household containing a couple with at least one aged over the pension age
Other family, no <18	- Household containing another family grouping (e.g. a multi-generational family), where there is no child aged under 18
Other family, with <18	- Household containing another family grouping (e.g. a multi-generational family), with one or more child aged under 18
Multi family, no <18	- Household containing at least one member who is not related to the other members (e.g. a family who have live-in servants), where there is no child aged under 18
Multi family, with <18	- - Household containing at least one member who is not related to the other members (e.g. a family who have live-in servants), with one or more child aged under 18

Users who wish to amend these household types would have to edit the data creation code (see Section 6).

5.2 The labor module log file

The next output to examine is the output of the labor supply module. This produces a number of tables that show the impact of the reforms on total (formal sector) earnings, and the impact of this change in earnings on tax revenues, given the assumed labor supply elasticities and the average and marginal effective tax rates people face. Figure 17 shows part of the logfile **logfile_labor** (in this case `labor_practise.txt`).

The first column shows pre-reform gross income, the second, the change in formal sector employment income, and the final column, the change in tax revenues associated with the change in formal sector income. Results are shown by age group, education group, sex and for the population as a whole.

In order to get the total revenue effects of the reform including the labor supply response, one must add the “Change in Tax” shown in this log file to the change in revenues shown in the **logfile_revenue** logfile (the pre-reform revenue change).

Figure 17. Showing how the burden is borne across the distribution

labor_practise - Notepad			
File Edit Format View Help			
Labor supply effects of reforms by educ group for system 20100			
Educ Group	Gross Income	Change in Income	Change in Tax
None or Preschool	\$ 87,266,590,720	\$ -66,302,380	\$ -5,250,465
Primary School	\$ 634,931,249,152	\$ -541,788,352	\$ -61,872,012
Secondary School	\$ 643,106,406,400	\$ -638,275,840	\$ -102,205,576
Degree (inc. PhD)	\$ 670,410,866,688	\$ -1,440,441,984	\$ -380,247,520
Commercial/Professional	\$ 1,381,925,257,216	\$ -3,432,065,536	\$ -940,344,960
Labor supply effects of reforms by age group for system 20100			
Sex	Gross Income	Change in Income	Change in Tax
Male	\$ 2,326,901,293,056	\$ -4,276,269,824	\$ -1,062,716,736
Female	\$ 1,090,739,175,424	\$ -1,842,604,288	\$ -427,203,776
Labor supply effects of reforms by age group for system 20100			
Total	Gross Income	Change in Income	Change in Tax
Total	\$ 3,417,640,271,872	\$ -6,118,874,112	\$ -1,489,920,512

5.3 The consumer demand module log files

The consumer demand module produces a number of log files. The first is the distributional file (**logfiledemand_welfare**), which shows the impact of the reforms to indirect taxes only by decile group and by household type(changes due to direct taxes have to be added on). Figure 18 shows the first part of the output of this file. Separate tables show the gains/losses in cash and proportional terms, and using different variables construct decile groups and different equivalence scales. The second column shows the cash or proportional gains/losses when one accounts for the ability of consumers to substitute between goods. This is measured by the compensating variation of the reform. The third column shows the cash or proportion gains/losses assuming no behavioural response. In this instance, allowing for demand effects makes little impact on the size of the losses to households.

Note that neither column is directly comparable to the figures contained in the main **logfiledecile** outputs because some approximation is required when conducting consumer demand analysis. If one wants the most accurate no behavioural response figures then the results of **logfiledecile** should be used rather than those from the **logfiledemand_welfare**.

Figure 18. The distributional impact of reforms including consumer demand responses

Decile	totexp	with Behavior	without Behavior
Poorest	\$ 38,626	\$ -103	\$ -103
Decile 2	\$ 57,940	\$ -182	\$ -183
Decile 3	\$ 70,275	\$ -245	\$ -246
Decile 4	\$ 78,243	\$ -288	\$ -289
Decile 5	\$ 90,825	\$ -364	\$ -365
Decile 6	\$ 102,458	\$ -431	\$ -432
Decile 7	\$ 116,323	\$ -514	\$ -515
Decile 8	\$ 135,926	\$ -620	\$ -622
Decile 9	\$ 166,351	\$ -811	\$ -813
Richest	\$ 324,373	\$ -1,676	\$ -1,677
Average	\$ 118,121	\$ -524	\$ -524

Decile	totexp	with Behavior	without Behavior
Poorest	\$ 38,626	-0.27%	-0.27%
Decile 2	\$ 57,940	-0.31%	-0.32%
Decile 3	\$ 70,275	-0.35%	-0.35%
Decile 4	\$ 78,243	-0.37%	-0.37%
Decile 5	\$ 90,825	-0.40%	-0.40%
Decile 6	\$ 102,458	-0.42%	-0.42%
Decile 7	\$ 116,323	-0.44%	-0.44%
Decile 8	\$ 135,926	-0.46%	-0.46%
Decile 9	\$ 166,351	-0.49%	-0.49%
Richest	\$ 324,373	-0.52%	-0.52%
Average	\$ 118,121	-0.40%	-0.40%

The next logfile produced is **logfiledemand_spending** which shows the impact of the reforms on spending patterns. Note that, in this example, we have set the global **inc_aff_spend = 1**, which means that in the change in income resulting from the change in direct taxes (as well as the change in prices resulting from changes in indirect taxation) is allowed to impact upon spending patterns. If **inc_aff_spend = 2**, instead, then only the impact of changes in indirect taxes on spending patterns would be modelled. Figure 19 shows the output of this logfile.

Figure 19. The impact of reforms on spending patterns

Commodity Group	Original	Reform 20100	Reform 20101
1	26.886%	26.898%	27.176%
2	12.905%	12.897%	12.832%
3	0.642%	0.649%	0.641%
4	7.168%	7.177%	7.160%
5	21.608%	21.567%	21.495%
6	1.633%	1.638%	1.641%
7	7.318%	7.288%	7.216%
8	6.344%	6.378%	6.394%
9	3.195%	3.193%	3.188%
10	7.580%	7.597%	7.590%
11	4.087%	4.078%	4.033%
12	0.634%	0.640%	0.634%

A list of commodity groups can be found in Appendix D. The first column shows the predicted fraction of aggregate consumer spending going towards each commodity group prior to the reforms. The second column shows the predicted fraction of aggregate consumer spending going towards each commodity group after the first set of modelled reforms (system 20100), with the last column showing the same for the second set of modelled reforms.

The last logfile produced by this module shows revenue figures before and after consumer demand responses are accounted for (declared in `logfiledemand_revenue`, in this instance `revenue_demand_practise.txt`). The first table shows the revenue raised from the changes to indirect taxes after consumer demand effects are allowed for, with the second column showing the revenue raised if one assumes there is no behavioural response. Figure 20 shows this.

Figure 20. Impact of consumer demand changes on revenues from indirect tax changes

Revenue raised from the tax reforms - after behavior		
Tax	Reform System 20100	Reform System 20101
IVA	\$ 6,753,457,664	\$ 30,980,372,480
IEPS	\$ 2,778,789,888	\$ 3,584,357,888

Revenue raised from the tax reforms - before behavior		
Tax	Reform System 20100	Reform System 20101
IVA	\$ 10,888,586,240	\$ 38,020,403,200
IEPS	\$ 3,064,007,424	\$ 4,083,248,640

5.4 The 'less than full pass through of indirect tax changes' log files

Three output files are produced that show the results of the simulations when less than full pass through of indirect taxes are assumed. These follow the same format as the outputs for the regular "no behavioral response" simulations and users are invited to examine that section for understanding how the outputs are arranged. It is worth mentioning, however, a number of conceptual/methodological issues here.

First, that the analysis produced by this module is very much an approximation. In particular, our input data does not identify which workers work for firms or which individuals are owners/shareholders of companies who would be affected by a particular set of indirect tax changes. Instead we change all income from wages and salaries by a fixed proportion so that the total change in wages and salaries is equal to the fraction of the indirect tax change borne by workers, and all income from profits or dividends by another fixed proportion so that the total change in income from these sources is equal to the fraction of the indirect tax change borne by shareholders. Unless the relevant workers or shareholders are distributed significantly differently to workers and shareholders more generally, this is likely to give at least a rough idea how less-than-full pass through would affect the distributional and revenue impact of reforms.

Second, the columns for the gains/losses due to income tax and social security contributions are replaced with a single column showing the change in net income. This is because less than full pass through of indirect tax changes implies gross income is affected by the reforms. This means net income is affected by not only changes in tax payments but also changes in gross income, and it is therefore clearest to show the impact of these two changes together, rather than separately.

The log files produced are:

`logfiledecile_ind` (in this case `decile_ind_p75_w75_practise.txt`)

`logfilehhtype_ind` (in this case `hhtype_ind_p75_w75_practise.txt`)

logfile_revenue_ind (in this case revenue_ind_p75_w75_practise.txt)

6 Input data

This section describes, briefly, how to create the input datasets required for use in LATAX for Mexico. The model requires three input files: a household demographics file (household.dta), an individual demographics and income file (adult.dta), and a household expenditure file (consumption_monetary.dta). Data creation files used to create our 2008 input datasets from the 2008 ENIGH can be found at the LATAX website (see above for URL): these can be used as a basis for developing input files with different variables or for different years. There are four do files used to create the input dataset, and they need to be run in the following order:

create_processed_data.do	- this file processes the raw ENIGH files to create files containing information on demographics, expenditure by category and household net income by source. It also creates variables identifying formal-sector workers and splits expenditure into formal sector and informal sector purchases.
create_grossincome.do	- this file converts net income by source into a gross income by source, using information on formality status from “create_processed_data.do” to determine whether tax has been paid by a particular individual. It also calls “clave_ingresos.do” in order to label data correctly. Further information on how this do file works and what is required to create gross income can be found in Section 6.2.
create_LATAX_input_data.do	- is the final do file which creates the household, adult and expenditure files used in LATAX itself from the files created by the previous do files. By default, these outputted files are called “household.dta”, “adult.dta” and “consumption_monetary.dta”, which can then be renamed for use in the simulator (e.g. as Input_Data_hh for the household file in the default implementation of LATAX).

Section 6.1 provides more detail on what is required of the household input file, section 6.2 covers the adult input file, and section 6.3 covers the expenditure input file. Section 6.4 provides a brief explanation on how to amend the input data to account for under-reporting of income and expenditure in the ENIGH survey.

6.1 The household input file

The household input file contains information on household composition and type, calculated equivalence scales, a sample weight and information sampling, a geographic identifier, and information on household net total, monetary and non-monetary income and expenditure. A full list of variables can be found in table C1 of Appendix C, with those in **bold** required and those in *italics* optional. We now describe the main required variables.

The first set of variables are the household identifiers. By default, these are **folioviv** and **foliohog**, which identify the address and the household within that address, respectively. These are taken from the ENIGH hogares.dta file. If the user wishes to rename these variables, or create a single identifier they can do so, and then amend the global **identlist** in main_module.do of LATAX. For instance, if a single variable called “**ident**” were created that uniquely identified each household, this would be changed to:

global identlist “ident”

The variable **numren** identifies individuals and in the household file is set equal to 1, allowing the household and adult level files to be merged when required.

The **hhtype** variable is a household type indicator, created in create_processed_data.do which is used to produce summary distributional analysis of reforms.

Three equivalence scales are required (**eqsc_5030**, **eqsc_8050**, and **eqsc_100100**), and these cannot be renamed without changing the main LATAX code. It is not advised to change what scale each variable refers to (e.g. **eqsc_5030** should be used to store the value of the scale where a first adult is given a weight of 1, second and subsequent people aged 12 or over are given a weight of 0.5, and children aged 11 or under a weight of 0.3), but one can do so, if one wishes. These are calculated in create_LATAX_input_dta.do.

The variable **factor** is the grossing factor used to weight the sample so that it matches the national Mexican population. It is taken directly from the hogares.dta ENIGH file. The variable **ubica_geo** is a geographical identifier which is taken from the hogares.dta ENIGH file and is used in consumer demand simulation and estimation.

The variables **totexp**, **monexp**, **nmonexp**, **moninc_e**, **moninc** and **totinc** are calculated variables which record the annual expenditure and income of a household as a whole. These are calculated from expenditure on individual items and individual income sources for each person in a household. It is possible to have **moninc** and **totinc** calculated directly by LATAX rather than having them as part of the household input file. To do this, the global **netincometype** in main_module.do should be set to 1. Section 6.4 describes how to amend the income and expenditure data so as to (roughly) account for the under-recording of income and expenditure in the ENIGH survey.

6.2 The adult input file

This file contains information on the demographics and gross incomes of individuals (including children as well as adults, although only the latter are used in practise). A full list of variables can be found in table C2 of Appendix C, with all those in **bold** required and the income variables in *italics* optional in the sense that the number of income sources can be amended (provided that the relevant parameters code is the amended in LATAX). We now describe the main required variables.

As with the household file, household (by default, **folioviv** and **foliohog**) and **individual** (**numren**) identifiers are required, and are taken directly from the ENIGH data. The program also requires the variables **sex**, **age**, **education**, and **rel_head** (relationship to head of household), all of which can be created from the pobla ENIGH file.

The variable **empstat** states whether an individual is working or not, and is based on the ENIGH variable **trabajo** which records whether someone worked in the last month or not. The variable **formal_w** is an indicator of whether a person is employed formally (and therefore assumed to pay tax and social security contributions on all income on which such payments are liable) or informally (and therefore is assumed to evade all such payments). The default is for this to be based on ENIGH variables which records whether an individual has access to healthcare provided by a social security scheme, and whether that access is via their own employment. This was the best proxy that could be found after significant search so it is not advised to change this without serious thought.

Finally, the remaining variables record the gross income an individual has from each source. The default is for there to be 20 sources (`inc1-inc20`), the first 10 of which record formal sector income, and the next 10 which record informal sector income (incomes are allocated to each source in `create_grossincome.do`, with formality based on the **formal_w** variable). The user is able to amend the number of sources of income when creating their own input dataset, although doing this will require changes to the parameters files (to change the number of sources, and the tax treatment of each source).

An important thing to note is that ENIGH provides data not on gross income as is required by a tax simulator like LATAX, but instead net income. This means the data creation programs must calculate gross income from the reported net income. To do this we “backwards calculate” gross income given net income and the tax rates, thresholds and exemptions that were in place during the year in which the survey was collected. Hence, the code contained in `create_grossincome.do` creates gross income from the 2008 ENIGH based on the 2008 tax and social security contributions system. If one wishes to use more up to date information one will need to replace the following information in `create_grossincome.do`:

- The amount of each income source that is subject to taxes and the amount that is exempt with that from the relevant year (see code starting after heading “Income including exemptions”), which usually depends upon some multiple of the monthly or daily minimum wage. Note that these amount are *monthly* amounts;
- The minimum wages with those from the relevant year (see code starting after the heading “NOW MAKING SURE ALL FORMAL PAID MIN WAGE”). Note that this should be expressed as an *annual* amount;
- The tax rates, fixed tax amounts (i.e. amount of tax paid according to tax rates below the current tax rate), tax credits, tax thresholds, and social security contribution rates with those from the relevant year (see code starting “BACK TO THE GROSS INCOME CALCULATOR”). The first matrix is for public sector workers subject to ISSSTE, and the second matrix is for private sector works subject to IMSS).

Checking that gross income is being calculated correctly is likely to be one of the trickiest parts of creating a new input dataset for LATAX and should not be rushed. It is worthwhile working out a number of examples whereby one calculates for a given gross income, the net income implied by the tax system, and then using the data creation code to recover the gross income. This should reduce the likelihood of errors.

6.3 The expenditure files

A full list of variables can be found in table C3 of Appendix C. All required variables are in bold. The expenditure variables are optional in the sense that the number and ordering of items can be changed (provided that the parameters and main_module files are correctly amended), although it is recommended to maintain the defaults for ease of use.

This file contains information on monetary expenditures only, separately by whether an item is deemed to be purchased formally or informally. The default is for the first 30 items to be formal sector items, and the next 30 items to be informal sector items. Expenditure by a household on an item is classified as informal if the household reports purchasing it from a market, stall or street vendor (ENIGH variable **lug_com** = 1, 2 or 3).

LATAX requires annual expenditures, meaning that the input data from ENIGH (which records some items on a weekly basis, others on a monthly, and others on longer time period) must be converted to an annual equivalent.

The final variable in this file is the region variable **region** which allows for VAT rates to differ for certain households. By default, **region** is set equal to 0 for all households, but setting **region** = 1 means that a household will face the **BVAT** rates of VAT defined in the parameters files, rather than the standard **VAT** rate. This allows one, in principle, to model the reduced rate of VAT in certain border areas of Mexico (if municipality boundaries align well with these areas, for instance).

6.4 Accounting for missing income and expenditure

The ENIGH surveys significantly under-record expenditure (particularly on items like alcohol and tobacco) and incomes (particularly capital income) relative to those recorded in the Mexican National Accounts. For instance, as reported in Abramovsky et al (2011), in order to match National Accounts figures, monetary employment income would need to be multiplied by a factor of 1.62/1.1, on average, self employment income by 2.52/1.1, capital and rental income by 26.04/1.1, and transfer income by 1.30 (the taxable income source factors are divided by 1.1 as the average tax rate is assumed to be around 10%).

The default input dataset does not make any adjustments to account for this under-reporting, but we do provide input datasets that account for under-recording using fixed factors for each income source and broad expenditure category to match National Accounts (see the LATAX website for these).

To adjust gross income to match national accounts when creating your own dataset, towards the end of the create_grossincome.do file, lines should be added multiplying the relevant income sources by the relevant factor. As with the analysis produced in Abramovsky et al (2011), one may decide to have factors that vary across the income distribution, or that are related to the extent to which expenditure is under-recorded (or vice versa).

To adjust household net income to match national accounts when creating your own dataset, amend a block of code should be inserted just before

```
* create different categories  
gen double moninc_e=inc_emp*12
```

in create_LATAX_input_data.do that multiplies each individual income source by the relevant factor. For example,

```
replace inc_emp = inc_emp*1.6173/1.1
replace inc_semp = inc_semp*2.5191/1.1
replace inc_cap = inc_cap*26.0441/1.1
replace inc_tran = inc_tran*1.2948
replace inc_oemp = inc_oemp*1.6173/1.1
replace inc_other = inc_other
```

is the code inserted in the do file that creates the alternative input dataset with adjustment for under-recording that we have made available for download from the LATAX website.

To adjust monetary expenditure for under-recording in ENIGH, code should be inserted in create_processed_data.do just prior to when each monetary expenditure file is saved (separate files are saved and then appended for data on regular expenditures, weekly expenditures recorded in the expenditure diary, and expenditure on education). The relevant places can be found by searching for “INSERT EXPENDITURE ADJUSTMENT FACTORS” in the create_processed_data.do file downloadable from the LATAX website.

It is important to note that national accounts provides information on only 12 categories of spending rather than the 60 (30 formal, 30 informal) included in LATAX. This means one must amalgamate the LATAX spending categories into larger categories (by default **exptype**=-1 to 10) for which National Accounts data is available, and multiply by factors calculated for these larger categories. For instance, the code inserted at the end of the code creating information on education spending is:

```
gen exptype = .
replace exptype = 8 if clave>="E001" & clave<="E007"

replace gas_tri = gas_tri*2.83 if exptype==-1
replace gas_mon = gas_mon*2.83 if exptype==-1
replace gas_tri = gas_tri*3.68 if exptype==0
replace gas_mon = gas_mon*3.68 if exptype==0
replace gas_tri = gas_tri*11.87 if exptype==1
replace gas_mon = gas_mon*11.87 if exptype==1
replace gas_tri = gas_tri*1.34 if exptype==2
replace gas_mon = gas_mon*1.34 if exptype==2
replace gas_tri = gas_tri*7.65 if exptype==3
replace gas_mon = gas_mon*7.65 if exptype==3
```

```
replace gas_tri = gas_tri*2.55 if exptype==4
replace gas_mon = gas_mon*2.55 if exptype==4
replace gas_tri = gas_tri*4.69 if exptype==5
replace gas_mon = gas_mon*4.69 if exptype==5
replace gas_tri = gas_tri*3.55 if exptype==6
replace gas_mon = gas_mon*3.55 if exptype==6
replace gas_tri = gas_tri*5.24 if exptype==7
replace gas_mon = gas_mon*5.24 if exptype==7
replace gas_tri = gas_tri*1.08 if exptype==8
replace gas_mon = gas_mon*1.08 if exptype==8
replace gas_tri = gas_tri*1.91 if exptype==9
replace gas_mon = gas_mon*1.91 if exptype==9
replace gas_tri = gas_tri*3.46 if exptype==10
replace gas_mon = gas_mon*3.46 if exptype==10
```

A list of the National Accounts categories can be found in table C4 in Appendix C.

7 Appendices

These appendices provide details on the 2010 Mexican reforms being used as an example in this manual. They also provide details on input variables, output variables, and the calculation modules.

Appendix A

Table A.1. A description of the 2010 tax reforms

Item	Status-quo	2010 tax reform <i>proposed</i> by the Executive power	2010 tax reform <i>approved</i> by the Congress and implemented
1. Income tax: both personal and corporate (Impuesto sobre la Renta – ISR).	Top three marginal rates are 19.94%, 21.95% and 28%.	-Top three marginal rates increase to 21.36%, 23.52% and 30% in 2010, 2011, 2012, with a phased reduction to 28% in 2014. -Individuals earning up to 4 minimum wages are not affected. -The annual upper threshold of income band 3 (lower threshold of income band 4) decreases from 88,793.04 \$ (mex) to 79,964.16.	-Top three marginal rates increase to 21.36%, 23.52% and 30% in 2010, 2011, 2012, with a phased reduction to 28% in 2014. -Individuals earning up to 6 minimum wages are not affected.
2. VAT (Impuesto al Valor Agregado - IVA)	General rate of 15%, and 10% in border areas	--	General rate of 16%, and 11% in border areas
3. Excise duties (Impuesto especial sobre la producción y servicios – IEPS)			
3.a. Tobacco	160% rate	Additional flat-rate of 0.04 for each cigarette or 0.75 grams of snuff; to be increased to 0.10 by 2014.	Additional flat-rate of 0.04 for each cigarette or 0.75 grams of snuff; to be increased to 0.10 by 2014.
3.b. Beer	25% rate	28% rate	26.5% rate (temporary)
3.c. Lottery	20% rate	30% rate	30% rate
3.d. Drinks with alcohol content greater than 20% by volume	50% rate	Additional minimum charge per litre of 3 pesos	53% rate
3.e. Telecommunications	None	4% rate	3% rate, except for Internet connexions
4. New expenditure tax (<i>Contribucion para el Combate a la Pobreza</i>)	--	Introduction of a 2% expenditure tax on all goods and services (with the exception of the purchase of government licenses and donations to charity)	Rejected
5. Tax on cash deposits	2% rate of balance	3% rate of balance	3% rate of balance

Source: CEFP (2009f)

Appendix B

This appendix provides details on the `main_module.do` module and the parameters`X'.do parameters files, so that users know what each option / parameter is used for.

B.1. `main_module.do`

This interface main module contains user-edited instructions on: the directories in which the simulation code, input and output data, and parameter code can be found; the name and location of logfiles; the names of input and output datasets; the existing and reform systems to be used in the analysis; and runs the tax simulator. There are options for whether one wants to conduct the behavioral analysis.

In order to run LATAX, the first thing necessary is to run `main_module.do`. Then the user must type in the following commands:

`set_systems` followed by the (numerical) name of 3 systems (e.g. 2008 20100 20101)

`set_defaults`

`simulator` which runs the program (calls `loadprogs.do`)

The default options for running LATAX are contained in '`main_module.do`' and can be directly edited '`main_module.do`'. Options can also be changed using interactive commands typed into the Stata command prompt.

Editable defaults in `main_module.do`

taxsim - the directory of the LATAX do files

in - the directory that contains LATAX input data

out - the directory that LATAX output data will be written to

param - the directory containing the parameter definition do files

logfiledecile - the log file for cash and proportional tax changes by decile group

logfiledecile1 - the log file for proportion of change in tax revenues attributable to each decile

logfilehhtype - the log file for cash and proportional tax changes by household type

logfilerevenue - the log file for revenue changes from tax reforms

indata_hh - household data input file

indata_ad - individual and incomes data input file

indata_co - expenditure data input file

outdata_ad - individual level output file

outdata_hh - household level output file

sys1 - baseline system number (e.g. 2010)

sys2 - reform system 1 number (e.g. 2011)

sys3 - reform system 2 number (e.g. 2012)

disttype - set equal to 1 to perform analysis with all equivalence scales, 2 for just 100/80/50 scale

netincometype- set equal to 1 to use net income calculated using LATAX, 2 to use a net income measure directly recorded in the input data

global identlist – the household identifier variable or set of variables

There are then a set of globals which are used to determine the type of behavioral analysis to be performed:

labmod - To run labor supply analysis set equal to 1 (2 otherwise)

logfilelabor - the log file for output of labor supply module

consmo - To run the QUAIDs demand analysis set equal to 1

inc_aff_spend - Set equal to 1 for demand modelling to account for changes in net income due to direct tax changes, 2 to exclude this effect (2 is default)

indata_demand - input file containing parameters of demand system

indata_prices - input file containing prices

logfiledemand_welfare - the log file for the welfare effects of tax reforms

logfiledemand_spending - the log file for the changes in spending patterns following tax reforms

logfiledemand_revenue - the log file for revenues from tax reforms allowing for demand effects

NUMGOODSQUAIDS - the number of good categories in the demand system (12 is default)

NUMGOODSDEM - the list of more-disaggregated goods (i.e. those contained in the indata_co file) included in demand system (a number of goods such as cars and other ‘lumpy’ purchases may have to be excluded from the demand system)

goodslist[1-12] - lists of disaggregated goods (i.e. those contained in the indata_co file) in each of the demand system categories

goodslist13 - A list of the disaggregated goods excluded from the demand system

categslis[1-60] - states the demand system category for each disaggregated good

indir - To perform analysis with less-than-complete IVA pass-through set equal to 1

prop_prices - The extent of pass through (between 0 and 1)

prop_wages - The extent to which taxes not passed on in prices are born by wages (between 0 and 1)

Four globals (corresponding to logfiledecile, logfilehhtype, logfilerevenue and outdata_ad) are used to define output files for analysis when different assumptions about IVA and duties pass-through are made and should be changed for each scenario to avoid over-writing files.

Changing options using interactive Stata commands

If you do not want to change the default options in main_module.do, you can also change most (but not all) options using interactive Stata commands. The following lists the optional commands and shows how to use them. For instance typing

```
set_progdir "C:\LATAX\Do_Files"
```

changes the taxsim global (see above) to “C:\LATAX\Do_Files”. The rest of the commands are written more generically in the following style:

```
set_progdir “[directory for taxsim]”
```

Remember DO NOT type the [].

The following commands can be typed to change the remaining file directories:

```
set_inputdir “[directory for in]”
```

```
set_outputdir “[directory for out]”
```

```
set_paramdir “[directory for param]”
```

Input and output filenames can be changed using the following commands

set_consdatt “[filename for **indata_co**]”

set_addatt “[filename for **indata_ad**]”

set_hhdatt “[filename for **indata_hh**]”

set_adout “[filename for **outdata_ad**]”

set_hhout “[filename for **outdata_hh**]”

The following commands change the names of the outputted logfiles from the no-behavioral response simulations:

set_logdec “[filename for **logfiledecile**]”

set_logbur “[filename for **logfiledecile1**]”

set_loghh “[filename for **logfilehh**]”

set_logrev “[filename for **logfilerevenue**]”

The following commands change options relating to how to carry out analysis

set_disttype [value for **disttype**]

set_netincometype [value for **netincometype**]

The following commands change options for the labor module

set_labmod [value for **labmod**]

set_labout “[filename for **logfilelabor**]”

The following commands change options for the consumer demand module

set_consmmod [value for **consmmod**]

set_changewithinc [value for **inc_aff_spend**]

set_demandparam “[filename for **indata_demand**]”

set_demandprice “[filename for **indata_prices**]”

set_demandwelfare “[filename for **logfiledemand_welfare**]”

set_demandspending “[filename for **logfiledemand_spending**]”

set_demandrevenue “[filename for **logfiledemand_revenue**]”

Finally, the following commands change options for the module which allows incomplete pass through of indirect changes to consumer prices:

set_indir [value for **indir**]

set_propprices [value for **prop_prices**]

set_propwages [value for **prop_wages**]

set_adout_ind “[filename for **outdata_ad_ind**]”

set_logdec_ind “[filename for **logfiledecile_ind**]”

set_loghh_ind “[filename for **logfilehhtype_ind**]”

set_logrev_ind “[filename for **logfilerevenue_ind**]”

Any **global** not listed above (e.g. NUMGOODSQUAIDS) cannot be changed interactively but must be amended directly in `main_module.do`.

B.2. `params_X.do`

The user defines the parameters of the VAT (IVA), duties (IEPS), income tax (ISR) and Social Security tax systems in the parameter do files. When naming the do files they should always be of the form `params_X.do`, where X is the name of the particular system (and corresponds to `sys1`, `sys2` or `sys3` as defined in *interface.do*). Existing parameters modules should be used as templates and it is important that the names of scalars are not changed (although of course the values can be).

For indirect taxes the user needs to define the following scalars:

- | | |
|-------------------|--|
| NUMGOODS | - The number of expenditure categories in the input data |
| OVAT[1-NUMGOODS] | - The standard VAT rate applicable to a particular category in the year the input data was collected. This should not be changed when changing the parameters of a reform or baseline system but only when the data used is from a different year (e.g. 2010). |
| OBVAT[1-NUMGOODS] | - If different VAT rates are applied in different parts of the country this allows a second rate of VAT to be set. It records the second rate applicable to a particular category in the year the input data was collected. This should not be changed when changing the parameters of a reform or baseline system but only when the data used is from a different year. |

ODUTIES [1-NUMGOODS] - The duties rate applicable to a particular category in the year the input data was collected. This should not be changed when changing the parameters of a reform or baseline system but only when the data used is from a different year.

VAT[1-NUMGOODS] - The standard VAT rate applicable to a particular category in the tax system under consideration. This should be changed when changing the parameters of a reform or baseline system.

BVAT[1-NUMGOODS] - If different VAT rates are applied in different parts of the country this allows a second rate of VAT to be set rate. This scalar defines the second VAT rate applicable to a particular category in the tax system under consideration. This should be changed when changing the parameters of a reform or baseline system.

DUTIES [1-NUMGOODS] - The duties rate applicable to a particular category in the tax system under consideration. This should be changed when changing the parameters of a reform or baseline system.

For income taxes the user needs to define the following scalars:

NUMSOURCES - The number of income sources in the input data

DEDUCTSOCSEC - Set equal to 1 if social security contributions are deducted from taxable income as they are in El Salvador

EXEMPT[1-NUMSOURCES] - The cash amount of a particular source that is exempt from income tax. For sources of income on which tax is not being simulated (e.g. capital income) this should be set to an amount higher than the largest observed value for that source.

PEXEMPT[1-NUMSOURCES] - The proportion of a particular source that is exempt from income tax.

MPEXEMPT[1-NUMSOURCES] - The amount of a source above which no additional proportion is exempt from income tax.

NUMBANDS - The number of income tax bands

BAND[0-NUMBANDS] - The upper-limit of each income tax band. The upper limit of the higher band should be higher than the largest observed value for gross income.

RATE[1-NUMBANDS] - Income tax rates (0 – 1)

MAXRATE - Maximum average tax rate if one applies

JUMP[1-NUMBANDS] - The discrete jump in tax liability that occurs when one enters a new tax band. This allows one to model ‘entry fees’ to the tax system (such as the minimum \$57.14 payment under the 2010 El Salvador system)

- NUMCREDS - Number of bands of employment income subsidy.
- LCRED[0-NUMCREDS] - Upper-limit of each employment-income subsidy band. The 0 band is to ensure that no subsidy is given to those with no earned income.
- ACRED[1-NUMCREDS] - Amount of credit for those with an income in each subsidy band.

For social security the user needs to define the following scalars:

- NUMSOCSECSYSTEMS - The number of systems with different rates and/or bands in operation
- NUMSOCSEC[1-NUMSOCSECSYSTEMS]- The number of bands for each of the social security systems

For each of the social security systems $X = 1..[NUMSOCSECSYSTEMS]$ and for each of the bands under that system $Y = 1.. NUMSOCSEC[1-NUMSOCSECSYSTEMS]$, the following parameter are also defined

- LSOCSEC X_Y - The upper-limit of band Y for social security system X .
- SOCSEC X_Y - The contribution rate applying to band Y for social security system X .
- SSNUMSOURCES - Number of sources of income for social security purposes
- SSEXEMPT[1-SSNUMSOURCES]- -The cash amount of a particular source that is exempt from social security. For sources of income on which tax is not being simulated (e.g. capital income) this should be set to an amount higher than the largest observed value for that source.
- SSPEXEMPT[1-SSNUMSOURCES] - The proportion of a particular source that is exempt from social security.
- SSMPEXEMPT[1-SSNUMSOURCES] - The amount of a source above which no additional proportion is exempt from social security.

Appendix C

Table C.1 Household file

Variables	Definition	ENIGH 2008 file	Variable description
folioviv	Residence identifier		
foliohog	Household identifier		
numren	Individual identifier		This is just to be able to merge results from the adult file collapsed at the household level. There is only one observation for each household and numren=01 for all households.
residents	Number of household members	Hogares	
menores	Number of household members under 12 years old	Concen	
mayores	Number of adult household members		residentes-menores
eqsc_5030	Equivalence scale		$1+ (\text{mayores}-1)*(0.5) + \text{menores}*(0.3)$
eqsc_8050	Equivalence scale		$1+ (\text{mayores}-1)*(0.8) + \text{menores}*(0.5)$
eqsc_100100	Equivalence scale		$1+ (\text{mayores}-1)*(1) + \text{menores}*(1)$
factor	Sampling weights	Hogares	
estrato	Stratum	Hogares	
ubica_geo	Area Identifier	Hogares	
hhtype	Household type	Pobla08	See hhtype.dta above
numfam	Number of families in the household	Pobla08	See hhtype.dta above
totexp	Annual total expenditure (monetary and non-monetary) in Mexican \$ 2008	gastos, gastodiaros, gastoeduca, nomonataro, hogares	Sum of all expnum categories defined in consumption file below + non-monetary expenditure irent + (autocon+ reminkind+ transfer) defined in nomonet_agg.dta above
monexp	Annual monetary expenditure in Mexican \$ 2008	gastos, gastodiaros, gastoeduca	Sum of all expnum categories defined in consumption file below
nmonexp	Annual non-monetary expenditure in Mexican \$ 2008	nomonataro, hogares	irent + (autocon+ reminkind+ transfer) defined in irent.dta and nomonet_agg.dta above

totinc	Annual net total (current) income (monetary + non-monetary) in Mexican \$ 2008	ingresos , nomonetario, hogares	(inc_emp+inc_semp + inc_cap+ inc_tran+ inc_oemp+ inc_other+ irent+ autocon +reminkind +transfer) defined in ingresos_mon_net_agg.dta and nomonet_agg.dta above
moninc	Annual net monetary (current) income in Mexican \$ 2008	ingresos	(inc_emp+inc_semp + inc_cap+ inc_tran+ inc_oemp+ inc_other) see definition in ingresos_mon_net_agg.dta
moninc_e	Annual net employment monetary (current) income in Mexican \$ 2008	Ingresos	inc_emp, see definition in ingresos_mon_net_agg.dta

Table C.2 Adult file

Variables	Formal/ Informal	Description	Definition	ENIGH file	2008	Variable description
folioviv			Residence identifier			
foliohog			Household identifier			
numren			Individual identifier	pobla08		
age			Age	pobla08		edad
sex			Sex	pobla08		sexo
education			Highest qualification achieved	pobla08		n_instr161
empstat			Whether in work	pobla08		Equals 1 if trabajo=1 or verifica=1,2,3,4,5
formal_w			Whether covered by social security through work	pobla08, trabajos		0=informal; 1=formal IMMS; 2=formal ISSSTE; 3 formal PEMEX
inc1	formal	Gross annual income in Mexican \$ 2008	Wages and salaries	ingresos, pobla08, trabajos		clave=P001, P002,P003, P006, P011, P015
inc2	formal	Gross annual income in Mexican \$ 2008	Overtime	ingresos, pobla08, trabajos		clave= P004
inc3	formal	Gross annual	End-of-the-year bonus	ingresos, pobla08,		clave=P009

		income in Mexican \$ 2008		trabajos	
inc4	formal	Gross annual income in Mexican \$ 2008	Incentives, rewards and prizes	ingresos, pobla08, trabajos	clave=P005
inc5	formal	Gross annual income in Mexican \$ 2008	Holiday bonuses and cash allowances	ingresos, pobla08, trabajos	clave=P007
inc6	formal	Gross annual income in Mexican \$ 2008	Profit sharing	ingresos, pobla08, trabajos	clave=P008, P019
inc7	formal	Gross annual income in Mexican \$ 2008	Pensions	ingresos, pobla08, trabajos	clave=P032
Inc8	formal	Gross annual income in Mexican \$ 2008	Other labor income	ingresos, pobla08, trabajos	Clave=P013, P017, P018 (net of P019)
inc9	formal	Net annual income in Mexican \$ 2008	Capital	ingresos, pobla08, trabajos	Clave= P016, P023-P031, P047, P065
inc10	formal	Net annual income in Mexican \$ 2008	Other income	ingresos, pobla08, trabajos	inc_semp (clave=P067-P080)+ inc_tran (clave=P033-P045, P066) + inc_other (clave=P046) + inc_oemp (clave=P020-P022, P063, P064)
inc11	informal	Gross annual income in Mexican \$ 2008	Wages and salaries	ingresos, pobla08, trabajos	clave=P001, P002,P003, P006, P011, P015
inc12	informal	Gross annual income in Mexican \$	Overtime	ingresos, pobla08, trabajos	clave= P004

2008					
inc13	informal	Gross annual income in Mexican \$ 2008	End-of-the-year bonus	ingresos, pobla08, trabajos	clave=P009
inc14	informal	Gross annual income in Mexican \$ 2008	Incentives, rewards and prizes	ingresos, pobla08, trabajos	clave=P005
inc15	informal	Gross annual income in Mexican \$ 2008	Holiday bonuses and cash allowances	ingresos, pobla08, trabajos	clave=P007
inc16	informal	Gross annual income in Mexican \$ 2008	Profit sharing	ingresos, pobla08, trabajos	clave=P008, P019
inc17	informal	Gross annual income in Mexican \$ 2008	Pensions	ingresos, pobla08, trabajos	clave=P032
inc18	informal	Gross annual income in Mexican \$ 2008	Other labor income	ingresos, pobla08, trabajos	Clave=P013, P017, P018 (net of P019)
inc19	informal	Net annual income in Mexican \$ 2008	Capital		Clave= P016, P023-P031, P047, P065
inc20	informal	Net annual income in Mexican \$	Other income		inc_semp (clave=P067-P080)+ inc_tran (clave=P033-P045, P066) + inc_other (clave=P046) + inc_oemp (clave=P020-P022, P063, P064)

Table C.3 Consumption file

All expenditure categories 'exp' are annual figures in Mexican \$ 2008.

Variables	Formal/ Informal	Tax classification 2008	Definition	ENIGH 2008 file	Variable description	Demand system categories	
						Variable name	Description
folioviv			Residence identifier				
foliohog			Household identifier				
exp1	Formal	VAT exempted	Health services	Gastos	(Clave=J001, J004, J005, J007, J008, J013, J016, J017, J018, J036, J039, J062, J072) & lug_com>3	expnum9	Non-taxed health and education
exp2	Formal	VAT exempted	Education services	Gastoeduca	(Clave=E001-E008, E015, E017, T905) & lug_com>3	excluded	
exp3	Formal	VAT exempted	Lottery	Gastos	(Clave=E029) & lug_com>3	expnum11	Leisure and hotels
exp4	Formal	VAT exempted	Public, school transport and transport abroad	Gastodiario	(Clave=B001-B007, E013, M001, T902) & lug_com>3	expnum8	Non-taxed transport goods and services and petrol
exp5	Formal	VAT exempted	Transfers and other services	Gastos	(Clave=G002-G006, G011, N006, N007, N011-N016) & lug_com>3	none (excluded)	
exp6	Formal	VAT exempted	Leisure goods and services	Gastos	Clave=E020-E022, E026 & lug_com>3	expnum11	Leisure and hotels

exp7	Formal	VAT exempted	Household good and services	Gastos	Clave=G012-G019 & lug_com>3	expnum6	Non-taxed household goods, services and electronics
exp8	Formal	VAT zero rate	Food and drinks	Gastodiario	(Clave=A001-A068, A070, A072-A197, A203-A215, A218, A242) & lug_com>3	expnum1	Non-taxed food and drinks
exp9	Formal	VAT zero rate	Education goods	Gastos	(Clave= E014) & lug_com>3	expnum9	Non-taxed health and education
exp10	Formal	VAT zero rate	Household good and services	Gastos	(Clave= G007) & lug_com>3	expnum6	Non-taxed household goods, services and electronics
exp11	Formal	VAT zero rate	Leisure services	Gastos	(Clave= L029) & lug_com>3	expnum11	Leisure and hotels
exp12	Formal	VAT zero rate	Health-related goods and services	Gastos	(Clave=J009, J010, J014, J020-J035, J037, J038, J042, J044-J059, J063, J064, T910) & lug_com>3	expnum9	Non-taxed health and education
exp13	Formal	VAT zero rate	Health-related goods and services	Gastos	(Clave=J061) & lug_com>3	expnum10	Taxed personal goods and services (including taxed health; education; etc)
exp14	Formal	VAT zero rate*	Food and drinks	Gastodiario	(Clave=A069, A071, A216, A217) & lug_com>3	expnum1	Non-taxed food and drinks
exp15	Formal	VAT taxed	Food and drinks	Gastodiario	(Clave=A198-A202, A219-A222, A243-A247, T901) & lug_com>3	expnum2	Taxed food, drinks and food out
exp16	Formal	VAT taxed	Household goods and services	Gastos	(Clave= C001-C024, T903, G008-G010, G020-G022, T909, F007, T911, I001-I026, K001-K044, L001-L022, T907) & lug_com>3	expnum5	Taxed household goods, services, communications and electronics
exp17	Formal	VAT taxed	Personal goods and services, including education and health	Gastos	(Clave= D001-D026, T904, E016, E018, E019) & lug_com>3	expnum10	Taxed personal goods and services (including taxed health; education; etc)

exp18	Formal	VAT taxed	Other non-food goods and services	Gastos	(Clave= E009-E012, M005, N008-N010, T915) & lug_com>3	none (excluded)	
exp19	Formal	VAT taxed	Leisure services	Gastos	(Clave= E023-E025, E027, E028, E030-E033, T912, L023-L028, N003-N005) & lug_com>3	expnum11	Leisure and hotels
exp20	Formal	VAT taxed	Transport goods and services	Gastos	(Clave= F013-F017, T906, T913, M002-M004, M006-M018) & lug_com>3	expnum7	Taxed transport goods and services and petrol
exp21	Formal	VAT taxed	Clothing, footwear and accessories	Gastos	(Clave= H001-H136, T908) & lug_com>3	expnum4	Taxed clothing and footwear
exp22	Formal	VAT taxed	Other non-food goods and services	Gastos	(Clave= N001-N002, T914) & lug_com>3	expnum12	Other (taxed and non-taxed, such as financial services)
exp23	Formal	VAT taxed	Health-related goods and services	Gastos	(Clave= J002, J003, J006, J011, J012, J015, J019, J040, J041, J043, J060, J065-J071) & lug_com>3	expnum10	Taxed personal goods and services (including taxed health; education; etc)
exp24	Formal	VAT taxed	Petrol	Gastos	(Clave= F010, F011, F012) & lug_com>3	expnum7	Taxed transport goods and services and petrol
exp25	Formal	VAT taxed	Telecoms	Gastos	(Clave= F001, F002, F003, F004, F005, F006, F008, F009) & lug_com>3	expnum5	Taxed household goods, services, communications and electronics
exp26	Formal	VAT and duties	Under 14° alcohol	Gastodiario	(Clave=A228, A231, A234, A232, A238) & lug_com>3	expnum3	Taxed Alcohol and tobacco
exp27	Formal	VAT and duties	14°-20° alcohol	Gastodiario	(Clave= A226, A237) & lug_com>3	expnum3	Taxed Alcohol and tobacco
exp28	Formal	VAT and duties	Over 20° alcohol	Gastodiario	(Clave = A223, A225, A227, A229, A230, A233, A235, A236) & lug_com>3	expnum3	Taxed Alcohol and tobacco
exp29	Formal	VAT and duties	Beer	Gastodiario	(Clave= A224) & lug_com>3	expnum3	Taxed Alcohol and tobacco

exp30	Formal	VAT and duties	Tobacco	Gastodiario	(Clave= A239, A240) & lug_com>3	expnum3	Taxed Alcohol and tobacco
exp31	Informal	VAT exempted	Health services	Gastos	(Clave=J001, J004, J005, J007, J008, J013, J016, J017, J018, J036, J039, J062, J072) & lug_com<4	expnum9	Non-taxed health and education
exp32	Informal	VAT exempted	Education services	Gastoeduca	(Clave=E001-E008, E015, E017, T905) & lug_com<4	excluded	
exp33	Informal	VAT exempted	Lottery	Gastos	(Clave=E029) & lug_com<4	expnum11	Leisure and hotels
exp34	Informal	VAT exempted	Public, school transport and transport abroad	Gastodiario	(Clave=B001-B007, E013, M001, T902) & lug_com<4	expnum8	Non-taxed transport goods and services and petrol
exp35	Informal	VAT exempted	Transfers and other services	Gastos	(Clave=G002-G006, G011, N006, N007, N011-N016) & lug_com<4	none (excluded)	
exp36	Informal	VAT exempted	Leisure goods and services	Gastos	Clave=E020-E022, E026 & lug_com<4	expnum11	Leisure and hotels
exp37	Informal	VAT exempted	Household good and services	Gastos	Clave=G012-G019 & lug_com<4	expnum6	Non-taxed household goods, services and electronics
exp38	Informal	VAT zero rate	Food and drinks	Gastodiario	(Clave=A001-A068, A070, A072-A197, A203-A215, A218, A242) & lug_com<4	expnum1	Non-taxed food and drinks
exp39	Informal	VAT zero rate	Education goods	Gastos	(Clave= E014) & lug_com<4	expnum9	Non-taxed health and education
exp40	Informal	VAT zero rate	Household good and services	Gastos	(Clave= G007) & lug_com<4	expnum6	Non-taxed household goods, services and electronics
exp41	Informal	VAT zero rate	Leisure services	Gastos	(Clave= L029) & lug_com<4	expnum11	Leisure and hotels

exp42	Informal	VAT zero rate	Health-related goods and services	Gastos	(Clave=J009, J010, J014, J020-J035, J037, J038, J042, J044-J059, J063, J064, T910) & lug_com<4	expnum9	Non-taxed health and education
exp43	Informal	VAT zero rate	Health-related goods and services	Gastos	(Clave=J061) & lug_com<4	expnum10	Taxed personal goods and services (including taxed health; education; etc)
exp44	Informal	VAT taxed	Food and drinks	Gastodiario	(Clave=A069, A071, A216, A217) & lug_com<4	expnum1	Non-taxed food and drinks
exp45	Informal	VAT taxed	Food and drinks	Gastodiario	(Clave=A198-A202, A219-A222, A243-A247, T901) & lug_com<4	expnum2	Taxed food, drinks and food out
exp46	Informal	VAT taxed	Household goods and services	Gastos	(Clave= C001-C024, T903, G008-G010, G020-G022, T909, F007, T911, I001-I026, K001-K044, L001-L022, T907) & lug_com<4	expnum5	Taxed household goods, services, communications and electronics
exp47	Informal	VAT taxed	Personal goods and services, including education and health	Gastos	(Clave= D001-D026, T904, E016, E018, E019) & lug_com<4	expnum10	Taxed personal goods and services (including taxed health; education; etc)
exp48	Informal	VAT taxed	Other non-food goods and services	Gastos	(Clave= E009-E012, M005, N008-N010, T915) & lug_com<4	none (excluded)	
exp49	Informal	VAT taxed	Leisure services	Gastos	(Clave= E023-E025, E027, E028, E030-E033, T912, L023-L028, N003-N005) & lug_com<4	expnum11	Leisure and hotels
exp50	Informal	VAT taxed	Transport goods and services	Gastos	(Clave= F013-F017, T906, T913, M002-M004, M006-M018) & lug_com<4	expnum7	Taxed transport goods and services and petrol
exp51	Informal	VAT taxed	Clothing, footwear and accessories	Gastos	(Clave= H001-H136, T908) & lug_com<4	expnum4	Taxed clothing and footwear

exp52	Informal	VAT taxed	Other non-food goods and services	Gastos	(Clave= N001-N002, T914) & lug_com<4	expnum12	Other (taxed and non-taxed, such as financial services)
exp53	Informal	VAT taxed	Health-related goods and services	Gastos	(Clave= J002, J003, J006, J011, J012, J015, J019, J040, J041, J043, J060, J065-J071) & lug_com<4	expnum10	Taxed personal goods and services (including taxed health; education; etc)
exp54	Informal	VAT taxed	Petrol	Gastos	(Clave= F010, F011, F012) & lug_com<4	expnum7	Taxed transport goods and services and petrol
exp55	Informal	VAT taxed	Telecoms	Gastos	(Clave= F001, F002, F003, F004, F005, F006, F008, F009) & lug_com<4	expnum5	Taxed household goods, services, communications and electronics
exp56	Informal	VAT and duties	under 14° alcohol	gastodiario	(Clave=A228, A231, A234, A232, A238) & lug_com<4	expnum3	Taxed Alcohol and tobacco
exp57	Informal	VAT and duties	14°-20° alcohol	Gastodiario	(Clave= A226, A237) & lug_com<4	expnum3	Taxed Alcohol and tobacco
exp58	Informal	VAT and duties	over 20° alcohol	Gastodiario	(Clave = A223, A225, A227, A229, A230, A233, A235, A236) & lug_com<4	expnum3	Taxed Alcohol and tobacco
exp59	Informal	VAT and duties	beer	Gastodiario	(Clave= A224) & lug_com<4	expnum3	Taxed Alcohol and tobacco
exp60	Informal	VAT and duties	tobacco	Gastodiario	(Clave= A239, A240) & lug_com<4	expnum3	Taxed Alcohol and tobacco

Note: * This category is assumed to be VAT zero rate for the purpose of integrating the consumer demand system in the simulator. See explanations to table C.1 below. This means that the estimates of IVA revenues are slightly smaller but the difference is negligible since these drink categories are very small in terms of total expenditure.

Table C.4 National Account Consumption Categories and Factors

Category Number	Description	Multiplication Factor
-1	Food	2.83
0	Non-alcoholic beverages	3.68
1	Alcohol and tobacco	11.87
2	Clothing and footwear	1.34
3	Housing, electricity, gas, water and other fuels	7.65
4	Furnishings, and household equipment	2.55
5	Health	4.69
6	Transport	3.55
7	Leisure and Culture	5.24
8	Education	1.08
9	Hotels, cafes and restaurants	1.91
10	Miscellaneous goods and services	3.46

Table C.5 Adult (Individual) Output File Variables

Variable Name	Description
folioviv	Residence identifier
foliohog	Household identifier
numren	Individual identifier
sex	Sex
age	Age
education	Education level
empstat	Employment status
inc1 to inc20	Gross income by source as in table C2
hhtype	Household type indicator
numfam	Number of families in household
residentes	Number of residents in household
mayores	Number of adult household members
menores	Number of household members under 12 years old
eqsc_5030	Equivalence scale
eqsc_8050	Equivalence scale
eqsc_100100	Equivalence scale
factor	Sample Weight
estrato	Strata Identifier
ubica_geo	Geographic Identifier

totexp	Annual total expenditure (monetary and non-monetary) in Mexican \$
monexp	Annual monetary expenditure in Mexican \$
nmonexp	Annual non-monetary expenditure in Mexican \$
moninc_e	Annual net total (current) income (monetary + non-monetary) in Mexican \$
moninc	Annual net monetary (current) income in Mexican \$
totinc	Annual net employment monetary (current) income in Mexican \$
socsec_[sys1,sys2,sys3]	Annual social security contributions for each system
socrate_[sys1,sys2,sys3]	Marginal social security contribution rate for each system
inctax_[sys1,sys2,sys3]	Annual income tax net of tax credits for each system
taxrate_[sys1,sys2,sys3]	Marginal income tax rate for each system
ieps_tot[sys1,sys2,sys3]	Household total IEPS paid for each system
iva_tot[sys1,sys2,sys3]	Household total IVA paid for each system
hh_change_inctax[sys2,sys3]	Change in income tax paid between base and each reform system
hh_change_socsec[sys2,sys3]	Change in social security contributions paid between base and each reform system
hh_change_ieps[sys2,sys3]	Change in IEPS paid between base and each reform system
hh_change_iva[sys2,sys3]	Change in IVA paid between base and each reform system
hh_change_tot[sys2,sys3]	Change in total tax paid between base and each reform system
hhnet[sys1,sys2,sys3]	Household net income under each system

Table C.6 Household Output File Variables

Variable Name	Description
folioviv	Residence identifier
foliohog	Household identifier
numren	Individual identifier
hhtype	Household type indicator
eqsc_5030	Equivalence scale
eqsc_8050	Equivalence scale
eqsc_100100	Equivalence scale
factor	Sample Weight
totexp	Annual total expenditure (monetary and non-monetary) in Mexican \$
monexp	Annual monetary expenditure in Mexican \$
nmonexp	Annual non-monetary expenditure in Mexican \$
moninc_e	Annual net total (current) income (monetary + non-monetary) in Mexican \$
moninc	Annual net monetary (current) income in Mexican \$
totinc	Annual net employment monetary (current) income in Mexican \$
hh_change_inctax[sys2,sys3]	Change in income tax paid between base and each reform system
hh_change_socsec[sys2,sys3]	Change in social security contributions paid between base and each reform system
hh_change_ieps[sys2,sys3]	Change in IEPS paid between base and each reform system

hh_change_iva[sys2,sys3]	Change in IVA paid between base and each reform system
hh_change_tot[sys2,sys3]	Change in total tax paid between base and each reform system
hhnet[sys1,sys2,sys3]	Household net income under each system

Appendix D

This appendix presents the Quadratic Almost Ideal Demand System (QUAIDS) used in the *quaid.do* module in detail. First, the theoretical basis of the model is explained, before moving on to discuss how we estimated the model.

D.1 Assumptions and welfare impacts

The impact of tax reforms on consumer spending patterns and welfare can be simulated using a built-in Quadratic Almost Ideal Demand System (QUAIDS). This is a useful demand system that allows the share of each type of good in total expenditure to vary in a flexible way with goods able to be luxuries (i.e. having an income elasticity of greater than 1) at one level of total expenditure and necessities (i.e. having an income elasticity of less than 1) at another (Banks, Blundell and Lewbel (1997)). The model assumes that the utility obtained from any particular good is not affected by the amount one works (and therefore demand for goods is also unaffected), and it does not allow for positive or negative externalities from expenditure on certain goods (for instance fuel, alcohol and tobacco).

The Quadratic Almost Ideal Demand System (QUAIDS) is a generalisation of the Almost Ideal Demand System (AIDS) and is based on the following indirect utility function:

$$\ln V = \left\{ \left[\frac{\ln x - \ln a(p)}{b(p)} \right]^{-1} + \lambda(p) \right\}^{-1}$$

Where x is expenditure, $a(p)$, $b(p)$ and $\lambda(p)$ are defined as:

$$\ln a(p) = \alpha_0 + \sum_i \alpha_i \ln(p_i) + \frac{1}{2} \sum_i \sum_j \gamma_{ij} \ln(p_i) \ln(p_j)$$

$$b(p) = \prod_{i=1}^n p_i^{\beta_i}$$

$$\ln \lambda(p) = \sum_{i=1}^n \lambda_i \ln(p_i)$$

where ($i=1, \dots, n$ denotes a good). Applying Roy's identity this gives the following equation for w_i , the share of expenditure on good i in total expenditure is, for each household:

$$w_i = \alpha_i + \sum_{j=1}^n \gamma_{ij} \ln(p_j) + \beta_i \ln\left(\frac{x}{a(p)}\right) + \frac{\lambda_i}{b(p)} \left(\ln\left(\frac{x}{a(p)}\right) \right)^2$$

For the resulting demands to be consistent with utility maximisation, the demand system must satisfy four key properties: adding-up; homogeneity; symmetry; and negativity (negative semi-definiteness). The first three can be imposed using linear restrictions on the parameters of the model:

(adding up)

$$\sum_{i=1}^n \alpha_i = 1; \quad \sum_{i=1}^n \beta_i = 0; \quad \sum_{i=1}^n \gamma_{ij} = 0 \forall j \quad \sum_{i=1}^n \lambda_i = 0$$

(homogeneity)

$$\sum_{j=1}^n \gamma_{ij} = 0 \forall i$$

(symmetry)

$$\gamma_{ij} = \gamma_{ji}$$

Negativity cannot be imposed in such a manner but the estimated Slutsky matrix can be tested to see if it satisfies this criterion.

This paper allows for household demographics to affect demands in a fully theoretically consistent manner. Demographics enter as taste-shifters in the share equations, and to maintain integrability they are therefore part of α_i terms in $\ln a(p)$:

$$\ln a(p) = \alpha_0 + \sum_i \left\{ \alpha_i + \sum_{k=1}^K \alpha_{ik} z_k \right\} \ln(p_i) + \frac{1}{2} \sum_i \sum_j \gamma_{ij} \ln(p_i) \ln(p_j)$$

Which gives us the following new adding-up conditions that supersede $\sum_{i=1}^n \alpha_i = 1$:

$$\sum_{i=1}^n \alpha_i = 1; \quad \sum_{i=1}^n \alpha_{ik} = 0;$$

Calculating the Welfare Impact of Price Changes

Having estimated a fully specified demand system, one can estimate the impact of price changes on consumer welfare using the associated expenditure functions. An attractive measure of the welfare impact is the compensating variation (CV): the change in income a household would require in order to make them indifferent between the original price vector (with the original income) and the new price vector. This is calculated as

$$CV = E(u^*, p^1) - E(u^*, p^0)$$

where u^* is the original value of the utility index, p^0 is the initial price vector, p^1 is the new price vector and $E(u^*, p^y)$ ($y=0,1$) is

$$E(u^*, p^y) = e^{\ln a(p^y) + b(p^y) \left\{ \frac{1}{\ln u^*} \lambda(p^y) \right\}^{-1}}$$

and where $\ln u^*$ can be calculated using the indirect utility function. Price and total expenditure elasticities are derived and presented in Banks et al (1997).

D.2 Econometric and empirical specification

The QUAIDS is estimated using a 2-step procedure programmed in STATA, with standard errors calculated using a clustered bootstrap procedure. Because total expenditure may be endogenous we instrument for it using monetary income. This is done using a control function approach.¹⁰

Stage 1

Before estimation, $a(p)$ and $b(p)$ are unknown. For this reason, $\ln a(p)$ is approximated using the Stone price index

$$\ln p^* \approx \sum_i w_i \ln p_i$$

and $b(p)$ is approximated as 1. Conditional upon the price indices, QUAIDS is linear in parameters. Hence, a linear Seemingly Unrelated Regression (SURE) framework is used to estimate the model. Adding up is imposed by excluding the equation for the n th good from the estimated system of equations; parameters for this equation are calculated using the parameters from the other $(n-1)$ equations and the adding up restrictions. Homogeneity and symmetry are imposed using linear restrictions on parameters.

Stage 2

The parameters estimated in the first stage are used to calculate values for $a(p)$ and $b(p)$. The model is then re-estimated using the same specification as the first stage except that p^* is replaced with $a(p)$ and λ_i by $\frac{\lambda_i}{b(p)}$. The new parameter values used to update $a(p)$ and $b(p)$, and the model is then re-estimated for a third time. This updating of price indices and re-estimation is iterated 12 times, by which point the parameter values have converged to 5 decimal places.

Standard errors are calculated using bootstrapping with 500 iterations. Rather than draw the bootstrap samples in an unrestricted manner we take into account that we use variation in prices across city-regions clusters and draw, with replacement, from within clusters as opposed to from the entire sample.

D.3 Data description

This section provides further information about the data used in the estimation of the demand system. These data should be the same as the one used in the MEXTAX micro-simulator. To generate these data the program `create_demand_data_dems.do` has been used. The demand system is estimated in the program `'quaid estimation_dems.do'` which is downloadable from the LATAX website (see above).

In order to ensure that the model can be feasibly estimated it is necessary to aggregate the very detailed expenditure breakdowns in ENIGH into a significantly smaller number of aggregate commodity groups. These are designed to ensure both that the groups make sense as functional

¹⁰ That is we regress $\ln x$ and $(\ln x)^2$ on the prices and demographic variables included in our demand system and on the log of household monetary income and the square of the log of household monetary income and include cubic terms of the residuals from these regressions in our demand system equations.

product groups but also to allow for substitution between goods treated differently by the indirect tax system. The 12 categories chosen are¹¹:

- Food and drinks on which no IVA is levied
- Food and drinks and meals out on which IVA is levied
- Alcoholic drinks and tobacco (IVA and IEPS levied)
- Clothing and footwear (IVA levied)
- Household goods, services and communications (IVA levied, IEPS sometimes levied)
- Household goods, services and communications (no IVA levied)
- Transport and vehicle fuels (IVA levied, IEPS sometimes levied but not modelled)
- Public transport and other transport on which no IVA levied
- Health and education goods (no IVA levied)
- Health and personal goods and services (IVA levied)
- Leisure and hotel services (IVA sometimes levied)
- Other services (IVA sometimes levied)

By aggregating goods in such a way, our demand model is suitable for modelling the welfare impacts of changing the rate of IVA and imposing IVA on additional classes of goods. However, this level of aggregation means that we cannot model, for instance, substitution between different kinds of alcoholic beverage when the duties rates on different types of beverages change by different amounts. Whilst this limits the number of questions the existing demand model can be used to assess, we would argue that analysis of very detailed goods categories is best done using bespoke demand systems tailored to the question at hand.¹²

Prices (and associated expenditure weights) used to calculate the prices of the aggregate commodities have been provided to us by the Bank of Mexico for 46 cities. The Bank of Mexico also provided data on the city whose prices are used for each municipality with a population of greater than 15,000 that is included in the ENIGH. The Bank determines these linkages using distance, population size, and other characteristics. Links between the cities and municipalities of less than 15,000 people are not made by the Bank and are instead computed by the authors using travel time according to Google maps. This simple method was chosen to ensure ease of replication by other researchers and for future (and past) waves of ENIGH.

The prices of the aggregated commodities are calculated as weighted arithmetic averages of the prices of the individual goods making up the commodity. Arithmetic as opposed to geometric averages (termed Stone prices) are used because geometric averages assume within-group own price elasticities of -1 and compensating cross-price elasticities which would mean that the welfare costs of changes in indirect taxes would be lower (than when not allowing for behavioral response) by assumption rather than because of the demand system estimates of the potential for substitution between different aggregate commodities.

We model changes in indirect taxes as increases in the prices of the aggregate commodities. When a commodity contains goods which are seeing differential proportional increases in prices due to changes in tax (for instance, the alcohol and tobacco group where different

¹¹ A full description of the products in each category can be found in table C.1 in Appendix C.

¹² For instance, if one wanted to estimate the impact of differential taxation of forms of alcohol and changes in alcohol taxation, one may want as categories the various forms of alcohol, tobacco, non-alcoholic drinks, food-out, food-in, other leisure, and "other goods and services".

changes may be made to the duty rates), the increase in the price of that commodity is the arithmetic weighted average of price changes of the goods within the category (averaged across the entire population). If the within-group composition differs systematically across the population (for instance, poorer households consuming alcohol in the form of beer, and richer households in the form of wine), or the extent of tax evasion varies across the population, this estimated change in prices might not reflect well the change in prices for the types of goods within that group that particular households face. This is a limitation of the demand modelling (more so for the analysis of changes in IEPS than for changes in IVA) and one that should be improved upon in future analysis of consumer responses to expenditure taxation in Mexico.

Our measure of the welfare cost of indirect tax changes is called the compensating variation (CV), that is the amount of income required to allow a consumer to obtain the same level of utility when prices increase (due to increases in indirect tax) once one allows for the ability to substitute between goods. We can then compare the averages of these CVs to averages of the welfare costs of indirect tax changes under the assumption of no behavioral response for broad groups of people (e.g. expenditure decile groups).

We use a demand model that has been estimated using unadjusted ENIGH data and this simulation has been conducted using the baseline (unadjusted) data for this reason. It is important that the estimation and simulation use the same data (e.g. the same corrections for the under-reporting of income or expenditure) otherwise the predictions of the demand system will not correspond well with the observed patterns of expenditure in the data used in the MEXTAX simulator.

Table D.1 lists and describes the 12 goods categories included in the demand system, providing detail of the ENIGH expenditure categories included, the Bank of Mexico Prices index codes, and the assumptions made when prices are unavailable for certain goods. Table D.2 lists and describes the demographic variables included in the demand systems.

Table D.1 Definition of categories used in the demand system

Categories	Variable name	ENIGH codes	Bank of Mexico PRICES codes	Notes
Food and drink on which no IVA is levied	expnum1	A001-A197, A203-A214 A215-A216, A217-A218, A242	1-100, 103-110	Prices for A215 (still bottled water - zero rate) and A216 (sparkling water -iva is levied on this) are together so include in category 1 as A215 represents a bigger expenditure share in ENIGH 2008. A218 (bottled juices or cordials -zero rate) and A217 (prepared water and natural juices -iva is levied on this) are together, include in 1 as A218 represents a bigger expenditure share in ENIGH 2008. No specific price for A242 (food dispensed by government agencies or NGOs).
Food and drink and meals out on which IVA is levied	expnum2	A198-A202, A219-A222, A243-A247, T901	101-102, 111-115, 378-381	
Alcoholic Drinks and Tobacco (IVA and IEPS levied)	expnum3	A223-A241	116-122	
Clothing and footwear (IVA levied)	expnum4	clave>="H001" & clave<="H136", clave=="T908"	134-171	No prices for jewelry, wrist-watches and other women accessories (Banxico code 172, ENIGH code H125-H127, H129, H131): include them and use price of 171 (handbags)

Household goods, services and communications (IVA levied, IEPS sometimes levied)	expnum5	clave>="C001" & clave<="C024", G009-G010, G020-G022, I001-I026, K001-K044, T903, T907, T909, T911, electronics (L001-L022), F001-F009, G008.	184, 186, 189, 190-195, 205-246, 294, 355, 361	No information on the price level of electricity (Banxico code 188, ENIGH code G008); use the price of gas (Banxico code 189). No information on the price level of landline telephone services (Banxico code 190 and ENIGH code F003); or national inter-city phone calls and international calls (Banxico codes 191 & 192 and ENIGH code F002); use the price of internet connexion (Banxico code 355). No price information on mobile telephone services (ENIGH codes F004 and F005); public telephones (ENIGH codes F006); and mail and other communications services (ENIGH codes F007, F009).
Household goods, services and communications (no IVA levied)	expnum6	G007, G011-G019	185, 187	Exclude housing renting cost and imputed rent from this category (Banxico codes 182 and 183 and ENIGH codes G001-G006)
Transport and vehicle fuels (IVA levied, IEPS sometimes levied but not modelled)	expnum7	F010-F017, M002-M004, M006, M012-M018, T906, T913,	310-318, 320-323, 325	No price level information for tolls (Banxico code and ENIGH code M005) or car insurance (Banxico code 319 and ENIGH code N008) ; exclude them. Exclude purchases of automobiles (including cars, bicycles, etc, Banxico codes 311 and 312, ENIGH codes M007-M011)
Public Transport and other transport on which no IVA levied	expnum8	B001-B007, M001, E013, T902	305-309	
Health and Education goods (no IVA levied)	expnum9	Health (J001, J004, J005, J007, J008, J013, J016, J017, J018, J036, J039, J062, J072, J009, J010, J014, J020-J035, J037, J038, J042, J044-J059, J063, J064, T910); Education (E014)	256-265, 268, 269, 271, 272, 276,277, 342, 343	No price level information for education fees (Banxico codes 335-342; ENIGH codes E001-E008, E017, E015, T905), exclude them from the system.
Health and personal goods and services (IVA levied)	expnum10	Taxed education (E016, E018, E019); Taxed health (J002, J003, J006, J011, J012, J015, J019, J040, J041, J043, J060-J061, J065-J071); Personal goods and services (D001-D026, T904)	266-267, 270, 273-275, 278-293, 295, 344-346	Prices for J060 (cotton wool, dressings - iva levied on this) and J061 (surgical alcohol -zero rated) are together so include in category 10 as J060 represents a bigger expenditure share in ENIGH 2008. Exclude education services E009-E012.

Leisure and hotel services (IVA sometimes levied)	expnum11	Non-taxed leisure (E020-E022, E026, E029, L029); Taxed leisure (E023-E025, E027-E028, E030-E033, L023-L028, N003-N005, T912)	347, 349-354, 356-360, 362-364,	No specific price for lottery (E029) or hotels.
Other services	expnum12	N001-N002, T914	382-384	No prices information for other services ENIGH codes N006, N007, N009-N010; exclude them from the analysis.

Table D.2 Other variables used in the demand system

Variable name	Description
<i>Demographic variables</i>	
child	Number of household members under 12 years old
adults	Number of household members 12 years or over
sex	= 1 if the head of the household is female, 0 otherwise
empstat	= 1 if the head of the household is employed, 0 otherwise
educlow	= 1 if the head of the household has primary education or less, 0 otherwise
educmid	= 1 if the head of the household has secondary education, 0 otherwise
central	Households in municipalities associated with the following cities: Cuernavaca, Puebla, Queretaro, Guadalajara, Aguascalientes, San Luis Petosi, Tlaxcala, Morelia, Jacona, Iguala, Leon, Tepatitlan, Toluca, Tulancingo, Cortazar
north_interior	Households in municipalities associated with the following cities: Jimenez, Ciudad Juarez, Chihuahua, Monclova, Monterrey, Torreon, Durango, Fresnillo, Ciudad Acuna
north_coastal	Households in municipalities associated with the following cities: La Paz, Mexicali, Matamoros, Hermosillo, Huatabampo, Tijuana, Culiacan
west	Households in municipalities associated with the following cities: Tepic, Acapulco, Colima
east	Households in municipalities associated with the following cities: Cordoba, Veracruz, San Andres Tuxtla, Tampico
south	Households in municipalities associated with the following cities: Oaxaca, Tehuantepec, Tapachula, Villahermosa
south_east	Households in municipalities associated with the following cities: Campeche, Chetumal, Merida

D.4 Estimated elasticities

Table D.3 shows the estimated compensated price elasticities and table C.4 show the income elasticities. The own-price elasticities confirm that the model satisfies negativity.

Table D.3 Hicksian (Compensated) price elasticities

Good	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) Food on which no IVA is levied	-0.229	0.039	-0.040	0.051	0.087	-0.002	-0.013	0.069	0.015	0.041	-0.010	-0.009
(2) Food on which IVA is levied and meals out	0.077	-0.847	0.072	0.010	0.255	0.041	0.068	0.122	0.043	0.137	0.011	0.011
(3) Alcoholic Drinks and Tobacco (IVA and IEPS levied)	-1.762	1.390	-0.744	-0.370	-0.410	0.210	0.343	0.325	0.205	0.172	0.416	0.225
(4) Clothing and footwear (IVA levied)	0.181	0.043	-0.044	-1.049	0.530	0.070	0.125	0.026	0.018	0.063	0.052	-0.014
(5) Household goods, services and communications (IVA levied, IEPS sometimes levied)	0.199	0.191	-0.018	0.183	-0.886	0.006	0.027	0.083	0.029	0.108	0.059	0.020
(6) Household goods, services and communications (no IVA levied)	0.035	0.254	0.073	0.207	0.031	-1.050	0.065	0.312	-0.004	-0.043	0.168	-0.049
(7) Transport and vehicle fuels (IVA levied, IEPS sometimes levied but not modelled)	-0.180	0.246	0.058	0.166	0.184	0.041	-0.841	0.051	0.069	0.048	0.154	0.004
(8) Public Transport and other transport on which no IVA levied	0.364	0.286	0.041	0.037	0.223	0.108	0.041	-1.217	0.050	0.112	-0.036	-0.009
(9) Health and Education goods (no IVA levied)	0.204	0.189	0.050	0.030	0.166	-0.001	0.079	0.106	-0.954	0.012	0.085	0.034
(10) Health and personal goods and services (IVA levied)	0.177	0.276	0.018	0.058	0.269	-0.014	0.031	0.094	0.005	-0.934	0.019	0.001
(11) Leisure and hotel services (IVA sometimes levied)	-0.107	0.047	0.108	0.087	0.428	0.140	0.206	-0.075	0.102	0.046	-1.038	0.057
(12) Other services	-0.697	0.360	0.374	-0.214	0.821	-0.231	0.044	-0.133	0.238	0.005	0.366	-0.932

Notes: Standard errors have not yet been calculated (due to the time the bootstrapping process takes). An updated version of this table will be provided when this has been completed. Elasticities are estimated using mean prices and expenditures and for a household with 2 adults and 2 children, where the head is male, has low levels of education, is employed and lives in the DF.

Source: Authors' calculations using MEXTAX, Bank of Mexico price data and ENIGH 2008.

Table D.4 Income elasticities

Good	Income Elasticity
(1) Food on which no IVA is levied	0.52
(2) Food on which IVA is levied and meals out	1.34
(3) Alcoholic Drinks and Tobacco (IVA and IEPS levied)	1.16
(4) Clothing and footwear (IVA levied)	1.20
(5) Household goods, services and communications (IVA levied, IEPS sometimes levied)	1.20
(6) Household goods, services and communications (no IVA levied)	0.84
(7) Transport and vehicle fuels (IVA levied, IEPS sometimes levied but not modelled)	2.06
(8) Public Transport and other transport on which no IVA levied	0.66
(9) Health and Education goods (no IVA levied)	1.12
(10) Health and personal goods and services (IVA levied)	0.98
(11) Leisure and hotel services (IVA sometimes levied)	2.09
(12) Other services	1.69

Notes: Standard errors have not yet been calculated (due to the time the bootstrapping process takes). An updated version of this table will be provided when this has been completed. Elasticities are estimated using mean prices and expenditures and for a household with 2 adults and 2 children, where the head is male, has low levels of education, is employed and lives in the DF.

Source: Authors' calculations using MEXTAX, Bank of Mexico price data and ENIGH 2008.

The patterns of substitution and complementarity seem reasonable. Food on which IVA is not levied (1) is a substitute for food on which IVA is levied and meals out (2). (2) is a substitute for both (1) and alcohol and tobacco (3). Private (7) and public (8) transport are also substitutes. Clothing (4) is complementary with (2) and (3), possibly reflecting additional demand for clothing when one is visiting restaurants and bars and other venues where food and alcohol are served.

The income elasticities are also sensible. Food on which IVA is not levied (1) is a necessity whilst food on which IVA is levied and meals out (2) is a luxury. The other strong necessity is public transport, whilst private transport, leisure goods and services and other services are strong luxuries.

D.5 Using LATAx with a different consumer demand model

It is beyond the scope of this manual to provide instructions on estimating a new consumer demand model. However, there may be circumstances when one wishes to do so. For instance, the general demand model used here groups all food commodities together, but if one wished to model the application of IVA/VAT on some but not all food items, one might wish to use a model which has disaggregated categories of food. In principle it is possible to do this, but a number of changes would have to be made to LATAx in order to do this.

First, one would have to change the relevant parts of `main_module.do`:

- The global **indata_demand** would have to be updated to the name of the input file containing the parameters of the new demand system, and the global **indata_prices** would have to be updated to the name of the input file containing the baseline prices for the goods in the new demand system.
- If the number of categories included in the demand system is going to be less than or greater than 12, **NUMGOODSQUAIDS** must be updated accordingly.
- If the new demand system requires disaggregating the main expenditure categories used in the no-behavioral-response simulations, the data creation programs must be amended accordingly and a new version of **indata_co** created. This will necessitate updating the parameters'X'.do files.
- The main expenditure categories used in the demand system need to be inputted in **NUMGOODSDEM**, and the expenditure categories in each demand system category need to be defined in the **goodslists**. The final **goodslist** should list the expenditure categories not included in any demand system category.
- The **catcglists** should be updated so that each main expenditure category is associated with a demand system category.

Second, if one plans on amending the demographic variables included in the demand system¹³ or if one wishes to change the measure of total spending used in the demand system (the demand system currently uses monthly spending equivalised using the `eqsc_8050` scale)), additional

¹³ The default variables are named: `child` (no of children), `adult` (number of adults), `sex` (sex of household head), `empstat` (employment status of household head), `educlow` (household head has low education), `educid` (household head has medium education), `v1-v6` (polynomials of the errors of first-stage regression of total expenditure on total income), `central`, `north_interior`, `north_coastal`, `west`, `east`, `south`, and `south_east` (regional identifiers).

changes will need to be made to the quaido.do file. In particular, one will need to change the global **dem**s and search for “eqsc_8050*12” in quaido.do and edit appropriately.

Changing the demand model from the QUAIDS form would require a comprehensive edit of quaido.do that is beyond the scope of this manual (and should only be considered by advanced users).

Finally, it is worth noting some good practises when developing a demand system for use with LATAX (or any tax microsimulation model):

- Ensure that there is sufficient exogenous variation in prices to estimate the demand system. This is unlikely to be the case if demand system categories are very narrow and there are few observations.
- Align demand system categories with tax system treatments (in baseline and reform systems) so that one can properly model the substitution that is predicted to occur from the relevant reform to the tax system.

References

Abramovsky, L., O. Attanasio, C. Emmerson, and D. Phillips (2011), "The distributional impact of reforms to direct and indirect tax in Mexico: Analytical Report and Results", Institute for Fiscal Studies

IFS (in consortium) (2011), "A retrospective evaluation of the EU VAT system", Report prepared for the EU Commission

Meyer, B. and J. Sullivan (2003), "Measuring the Well-Being of the Poor Using Income and Consumption", *Journal of Human Resources* 38, S1180-1220.

Meyer, B. and J. Sullivan (2004). "The Effects of Welfare and Tax Reform: The Material Well-Being of Single Mothers in the 1980s and 1990s," *Journal of Public Economics*, 88, July, 1387-1420.

Meyer, B. and J. Sullivan (2008), "Changes in the Consumption, Income, and Well-Being of Single Mother Headed Families," *American Economic Review*, 98(5), December, 2221-2241.

Meyer, B. and J. Sullivan (2011), "Further Results on Measuring the Well-Being of the Poor Using Income and Consumption", *Canadian Journal of Economics*, 44(1), 52-87.

Poterba, J. (1989). "Lifetime Incidence and the distributional burden of excise taxes", *American Economic Review*, 79, pp325-30.