Trade Policy Modeling with GTAP-WiNDC

Thomas F. Rutherford
University of Wisconsin – Madison

Department of Agricultural and Applied Economics
University of Wisconsin, Madison

WiNDC Short Course

June 14, 2022

Motivation

Analysis of policies affecting markets in multiple countries requires both data and theory. The Global Trade Analysis Project (GTAP) consortium provides data, and the analyst confronts this data with a theoretical perspective. Despite some limitations in data coverage and quality, a key practical constraint lies in the informed translation of theoretical insights into quantitative policy evidence.

The Wisconsin National Data Consortium (WiNDC) is a research project which has been formed to provide the analytic community with a dataset and a collection of companion models which facilitate evidence-based economic research on the national and sub-national level. The tools are open-source and may be used for multisectoral, multi-household general equilibrium analysis. The most recently WiNDC dataset is calibrated to align with GTAP version 10. The core data incorporates (approximately) quintile disaggregation of households, all US states, 10 regions/countries representing international trade and 32 economic sectors for 2014.

Blaise Pascal

Je n'ai fait celle-ci plus longue que parce que je n'ai pas eu le loisir de la faire plus courte.

"**I would have written a shorter letter, but I did not have the time.**"

Provincial Letters: Letter XVI, 4 December, 1656.

Such statements have also been attributed to Mark Twain, T.S. Eliot, Cicero, and others besides, but this article at Quote Investigator concludes that Pascal's statement is likely the original source of the phrase. *Wikipedia.*

Outline

- Mathiesen's modeling framework
- MPSGE
- CES functions and application
- Simulation analysis with GAMS and EXCEL
- WiNDC origins and motivation
- Equations of the GTAP-WiNDC model
- Policy application with GTAP-WiNDC: tariff quotas

Mathiesen's Modeling Format

In a pioneering paper Mathiesen (1985) presented both a modeling format and a solution algorithm for partial and general economic equilibrium problems. His paper reported on computational experience from a series of small to medium sized problems taken from the literature on the computation of economic equilibria. The common characteristic of these models was the presence of weak inequalities and complementary slackness, e.g., a linear technology with alternative activities or various institutional constraints on prices. The algorithm computed equilibria by solving a sequence of linear complementarity problems. The iterative (outer) part of this algorithm is a Newton process. The inner part used Lemke's almost complementary pivoting algorithm.

Methods and Applications

Theoretical results for the performance of Mathiesen algorithm are only available only for the partial equilibrium cases, however computational experience with general equilibrium models has proven to be remarkable. Much of the success of the approach is due to the inventiveness of Michael Ferris and his students who developed the PATH solver for mixed complementarity problems.

Mathiesen addressed both partial and general economic equilibrium problems involving production and consumption and presented a unifying treatment within a modeling format. The GTAP-WiNDC model is a general (or Walrasian) equilibrium problem, hence our account of Mathiensen's approach will be restricted to that format.

Optimization and Equilibrium

In a conventional general equilibrium model, consumers and producers are maximizers, and the *equilibrium conditions for the model as a whole correspond to first order necessary conditions for each agent or sector together with adding-up (market clearance) conditions which determine equilibrium prices*. Following Scarf (1973) Mathiesen's key insight was that that the resulting model is a complementarity problem, i.e.,

$$(CP) \quad \text{find} \quad z \in R^n \quad \text{that solves} \quad F(z) \geq 0, z \geq 0 \text{ and } z^T F(z) = 0$$

The Canonical General Equilibrium Model

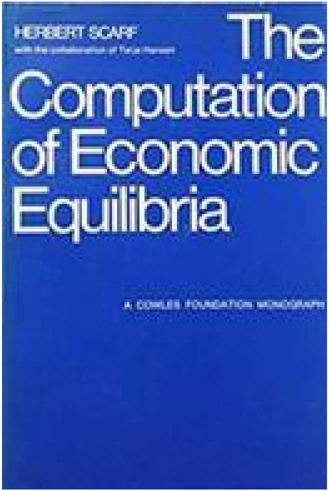
Consider an economy or a sector with production. Suppose that it has m commodities and n activities with *constant returns to scale* production. We base our presentation of the modeling format and the algorithm on the assumption that production is characterized by a technology matrix with *price-responsive* input-output coefficients.

For $i = 1, \ldots, m$ and $j = 1, \ldots, n$, let:

- $p = (p_i)$ denote the vector of prices,
- $b = (b_i)$ denote the vector of endowments,
- $d(p) = (d_i(p))$ denote the market demand functions, which we assume to be point-to-point and continuously differentiable,
- $y = (y_j)$ denote the vector of activity levels, and finally, let
- $A = (a_{ij}(p))$ denote the technology matrix of price-responsive input-output coefficients consistent with unit production at prices p, where $a_{ij} > 0$ $(a_{ij} < 0)]$ denotes an output (input).

Scarf with Lensburg

Because of the generality of the theory of economic equilibrium, there are several ways to characterize an equilibrium. Mathiesen chose to follow Chapter 5 of Scarf's monograph



Economic Equilibrium

A price vector p^* and a vector of activity levels y^* constitute a competitive equilibrium if:

i. No activity earns a positive profit:

$$-A^T p^* \geq 0.$$

ii. No commodity is in excess demand:

$$b + Ay^* - d(p^*) \geq 0.$$

iii. No prices or activity levels are negative: $p^* \geq 0, y^* \geq 0$.

iv. An activity earning a deficit is not used and an operated activity has no loss:

$$\left(-A^T p^*\right)^T y^* = 0.$$

v. A commodity in excess supply has zero price, and a positive price implies market clearance;

$$(p^*)^T (b + Ay^* - d(p^*)) = 0.$$

Observations

- All prices are determined simultaneously and no single price will be exogeneously given. In this case, demands $d_i(p)$ for $i = 1, \ldots, m$ are functions of all prices in the economy, i.e., both product and factor prices.
- The input-output matrix and demand functions are specified in a manner consistent with producer profit and household utility maximization, that is, $a_{ij}(p)$ maximizes

  $$\sum_i p_i a_i,$$

  and

  $$d_i(p) = \sum_h x_i^h$$

  ,
  where x_i^h is the hth household's utility maximizing demand of commodity i.

Observations (cont.)

- Households' excess demands are given by $d(p) - b$. If the demands satisfy each individual household's budget and there is nonsatiation, then $p^T d(p) = p^T b$, and the demand functions $d(p)$ are homogeneous of degree 0 in prices.
- Furthermore, conditions (i) – (v) determine only relative prices. That is, if the vector p^* represents equilibrium prices, so does λp^* for any scalar $\lambda > 0$. Hence, we are free to normalize the prices.

The General Equilibrium with Taxes

Many policy-oriented general equilibrium models deal with tax distortions (tariffs, factor taxes, value-added taxes, etc.) It is helpful to see that these distortions are easily integrated into Mathiesen's model. In the process we will display a generalization of the format which explicitly describes the sources of individual household income. In the model with closed financial flows, taxes on production and consumption show up in budget constraints of the "household" agents who collects tax revenue.

Tax Revenue Accounting

In the extended specification we introduce vector represent tax revenue vector $\tau_{jh}(p)$ representing the value of tax revenues paid by sector j to household h *per unit activity*.

The unit profit level of sector j is then

$$\Pi_j = \sum_i p_i a_{ij} - \sum_h \tau_{jh},$$

in which the first term represent the value of net market sales and the second term representing the value of tax obligations.[1]

[1] Taxes payments τ are functions of market prices, typically of an advalorem format, e.g. $\tau_{jh} = \sum_i p_i t_{ijh} a_{ij}$, but the functions τ may represent any number of tax instruments. The key idea is that when a tax is applied, it enters the zero-profit condition of the associated sector and it shows up in the budget constraint of the agent levying the tax.

Income Accounting

- This generalized formulation incorporates income balance conditions for each household, and this requires an explicit representation of household endowments.
- Let ω_{hi} denote the endowment of commodities i by household h
- The aggregate endowment vector in Mathiensen's framework is then given by $b_i = \sum_h \omega_{hi}$.
- The household utility maximization problem then takes the form:

$$x_{ih}(p, M_h) \quad \text{solves} \quad \max_x U_h(x) \text{ subject to } \sum_i p_i x_i = M_h$$

Equilibrium with Taxes

Prices p_i and activity levels y_j and household incomes M_h constitute a competitive equilibrium if:

i. No activity earns a positive profit net of taxes:

$$-\sum_i a_{ih} p_i + \sum_h \tau_{jh} \geq 0 \quad \forall j.$$

ii. No commodity is in excess demand:

$$\sum_h \omega_{ih} + \sum_j a_{ij} y_j - \sum_h x_{ih} \geq 0 \quad \forall i$$

iii. No prices or activity levels are negative: $p_i \geq 0 \; \forall i, \quad y_j \geq 0 \; \forall j.$

iv. An activity earning a deficit is not used and an operated activity has no loss:

$$\left(-\sum_i a_{ij} p_i + \sum_h \tau_{jh} \right) y_j = 0 \quad \forall j$$

Equilibrium with Taxes (cont.)

i. A commodity in excess supply has zero price, and a positive price implies market clearance;

$$p_i \left(\sum_h \omega_{ih} + \sum_j a_{ij} y_j - \sum_h x_{ih} \right) = 0 \quad \forall i$$

ii. The value of household expenditure is determined by household endowment income and tax receipts:

$$M_h = \sum_i p_i \omega_{ih} + \sum_j \tau_{jh} y_j \quad \forall h$$

References

Lars Mathiesen, "Computational Experience in Solving Equilibrium Models by a Sequence of Linear Complementarity Problems", *Operations Research* 33(6) 1985.

Herbert Scarf (with Terje Hansen), *The Computation of Economic Equilibria*, Cowles Foundation Monograph, 1973.

Michael Ferris, `https://pages.cs.wisc.edu/ ferris/`

Scalar Example

```
$TITLE: KYIV2  MGE-MCP Version in Mathiesen's Format

$ONTEXT
               Production Sectors     Consumers

   Markets    |    X       Y       W       I
   ------------------------------------------------------
      PX      |  100             -100  |
      PY      |          100     -100  |
      PL      |  -25     -75           |   100
      PK      |  -75     -25           |   100
      PW      |                   200  |  -200
   ------------------------------------------------------

This version is intended to closely relate to the MPSGE
formulation at the cost of using the $macro operator
and adding some variables.

$OFFTEXT
```

Scalar Example Declarations

```
PARAMETERS
 tr        ad valorem tax for X sector inputs on a NET basis /0/
 tc        iceberg transportation (trade) cost on X on a NET basis /0/
 lbar      labor endowment /100/
 kbar      capital endowment /100/;

NONNEGATIVE VARIABLES

 W        activity level for utility or welfare
 X        activity level for X production
 Y        activity level for Y production

 PL       price of labor
 PK       price of capital
 PX       price of good X
 PY       price of good Y
 PW       price of welfare (expenditure function)

 I        income of the representative consumer;
```

Scalar Example Declarations (cont.)

```
EQUATIONS

 PRF_W   zero profit for welfare
 PRF_X   zero profit for sector X
 PRF_Y   zero profit for sector Y

 MKT_L   supply-demand balance for primary factor L
 MKT_K   supply-demand balance for primary factor K
 MKT_X   supply-demand balance for commodity X
 MKT_Y   supply-demand balance for commodity Y
 MKT_W   supply-demand balance for welfare

 INC_I   income balance;

MODEL KYIV2 /PRF_W.W, PRF_X.X, PRF_Y.Y,
             MKT_L.PL, MKT_K.PK, MKT_X.PX, MKT_Y.PY, MKT_W.PW,
             INC_I.I /;
```

Scalar Example Macro Functions

```
*       Declare compensated demand and supply functions here:

$macro  I_PL_X  0.25*(PK/PL)**0.75
$macro  I_PK_X  0.75*(PL/PK)**0.25
$macro  O_X_PX  1/(1+tc)
$macro  R_X     tr * (PL * I_PL_X + PK * I_PK_X)

$macro  I_PL_Y  0.75*(PK/PL)**0.25
$macro  I_PK_Y  0.25*(PL/PK)**0.75
$macro  O_Y_PY  1
$macro  I_PX_W  0.5*PX**0.5*PY**0.5/PX
$macro  I_PY_W  0.5*PX**0.5*PY**0.5/PY
$macro  O_W_PW  1
```

Scalar Example Equations Declarations

```
*       Zero profit inequalities

PRF_W..      PX * I_PX_W + PY * I_PY_W =E= PW * O_W_PW;

PRF_X..      PL * I_PL_X  + PK * I_PK_X + R_X =E= PX * O_X_PX;

PRF_Y..      PL * I_PL_Y  + PK * I_PK_Y =E= PY * O_Y_PY;

*       Market clearance inequalities

MKT_L..      LBAR =G= I_PL_X * X + I_PL_Y * Y;

MKT_K..      KBAR =G= I_PK_X * X + I_PK_Y * Y;

MKT_X..      X * O_X_PX =G= I_PX_W * W;

MKT_Y..      Y * O_Y_PY =G= I_PY_W * W;

MKT_W..      W    =G= I/PW;

*       Income balance equation

INC_I..      I =E= LBAR*PL + KBAR*PK + R_X * X;
```

Scalar Example Benchmark Replication

```
*       Set initial values of variables (.L notation after variable)

X.L=100; Y.L=100; W.L = 200; I.L=200;
PX.L=1; PY.L=1; PK.L=1; PL.L=1;

*       Chose a numeraire: price of U fixed (.FX) at 1

PW.FX = 1;

KYIV2.ITERLIM = 0;

SOLVE KYIV2 USING MCP;

*       Abort with an error message if we have a deviation greater
*       than 0.001:

ABORT$round(KYIV2.OBJVAL,3) "Benchmark replication fails.";
```

Outline

- Mathiesen's modeling framework
- MPSGE
- CES functions and application
- Simulation analysis with GAMS and EXCEL
- WiNDC origins and motivation
- Equations of the GTAP-WiNDC model
- Policy application with GTAP-WiNDC: tariff quotas

MPSGE: A Mathematical Programming System for General Equilibrium Analysis

- Model representation tool for a specific class of economic models: Arrow-Debreu general equilibrium.
- Incorporates facilities for automatic calibration of cost and expenditure functions
- Model input is *tabular* (non-algebraic) and follows the broad schematic structure of the MPS format for general equilibrium models
- Provides routines for providing demand and supply functions (price-responsive netputs for production sectors and excess demands for consumers) and *analytic Jacobians*.
- Provides background error checks related to model consistency

MPSGE Historical Timeline

1981 Lars Mathiesen spends sabbatical at the Stanford OR Department with the research objective of implementing Newton's method for generalized equations [Josephy (1979)]

1982 A pilot implementation of the nonlinear complementarity solver MILES is completed, based in Tomlin's LCPL code for Lemke's algorithm and Saunders' LUSOL (the sparse matrix factorization code from MINOS)

1982 A trade policy research project, *Market Prospects*, is undertaken at NHH in Bergen. A central element of the project is a global Heckscher-Ohlin trade model, VEMOD. Project participants included Victor Norman, Agnar Sandmo, Lars Mathiesen, Terje Hansen, Terje Lensburg, and Erling Stigum.

1983 A standardized set of routines for representing nested CES functions is implemented to help with the ongoing formulation and reformulation of VEMOD.

1984 A pilot implementation of MPSGE is presented at TIMS XXVI, June 17-21, 1984 in Copenhagen, Denmark (seminar audience: 3)

The Impetus for GAMS (Alex Meeraus)

"GAMS's impetus for development arose from the frustrating experience of a large economic modeling group at the World Bank. In hindsight, one may call it a historic accident that in the 1970s mathematical economists and statisticians were assembled to address problems of development. They used the best techniques available at that time to solve multi sectoral economy-wide models and large simulation and optimization models in agriculture, steel, fertilizer, power, water use, and other sectors. Although the group produced impressive research, initial success was difficult to reproduce outside their well functioning research environment. The existing techniques to construct, manipulate, and solve such models required several manual, time-consuming, and error-prone translations into different, problem-specific representations required by each solution method."

GAMS Historical Timeline

1976 GAMS idea is presented at the ISMP Budapest

1978 Phase I: GAMS supports linear programming. Supported platforms: Mainframes and Unix Workstations

1979 Phase II: GAMS supports nonlinear programming.

1987 GAMS becomes a commercial product

1988 First PC System (16 bit)

1989 GAMS begins to be used as a front-end and back-end to MPSGE, producing input data matrices and model reports.

1991 Alex Meeraus collaborates on implementation of MPSGE and MCP as GAMS subsystems

1994 GAMS supports mixed complementarity problems

MPSGE Representation for the Scalar Model

```
$ONTEXT
$MODEL:KYIV2

$SECTORS:
        X       ! Activity level for sector X
        Y       ! Activity level for sector Y
        W       ! Activity level for sector W (Hicksian welfare index)

$COMMODITIES:
        PX      ! Price index for commodity X
        PY      ! Price index for commodity Y
        PL      ! Price index for primary factor L
        PK      ! Price index for primary factor K
        PW      ! Price index for welfare (expenditure function)

$CONSUMERS:
        CONS    ! Income level for consumer CONS
```

MPSGE Representation (cont.)

```
$PROD:X s:1
        O:PX  Q:(100/(1+tc))
        I:PL  Q:25   A:CONS T:tr
        I:PK  Q:75   A:CONS T:tr

$PROD:Y s:1
        O:PY  Q:100
        I:PL  Q:75
        I:PK  Q:25

$PROD:W s:1
        O:PW   Q:200
        I:PX   Q:100
        I:PY   Q:100

$DEMAND:CONS
        D:PW   Q:200
        E:PL   Q:LBAR
        E:PK   Q:KBAR
$OFFTEXT
$SYSINCLUDE mpsgeset KYIV2

$INCLUDE KYIV2.GEN
SOLVE KYIV2 USING MCP;
```

Outline

- Mathiesen's modeling framework
- MPSGE
- CES functions
- Simulation analysis with GAMS and EXCEL
- WiNDC origins and motivation
- Equations of the GTAP-WiNDC model
- Policy application with GTAP-WiNDC: tariff quotas

CES Technology

A constant-elasticity-substitution production function can be defined as:

$$y = f(x) = \left(\sum_i a_i x_i^\rho \right)^{1/\rho}$$

where $a_i > 0 \quad \forall i$

The CES production function may alternatively be written as:

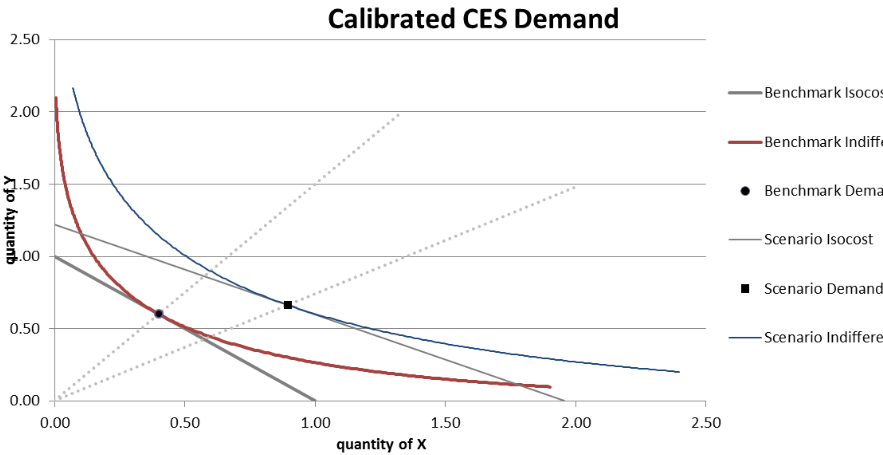
$$f(x) = \phi \left(\sum_i \alpha_i x_i^\rho \right)^{1/\rho}$$

where $\phi > 0$, $\alpha_i > 0$ and $\sum_i \alpha_i = 1$.
or

$$f(x) = \bar{y} \left(\sum_i \theta_i \left(\frac{x_i}{\bar{x}_i} \right)^\rho \right)^{1/\rho}$$

where $\theta_i \geq 0$ and $\sum_i \theta_i = 1$.

CES Demand



x0 0.40	px0 1.0	sigma 1.5	epsilon 2	px 0.625	py
Benchmark Value Share of X	Benchmark Price of X (py0=1)	Elasticity of Substitution	Price Elasticity of Aggregate Demand	Scenario Price of X	Scenario P

A Partial Model of Crude Oil Markets

- Regional supply:

$$Y_r = \bar{y}_r C_r^{\eta_r}$$

$\bar{y}_r$ Benchmark supply
η_r Price elasticity of supply

- Regional demand

$$D_r = \bar{d}_r P_r^{-|\epsilon_r|}$$

$\bar{d}_r$ Benchmark demand for products
ϵ_r Price elasticity of demand

A Partial Model of Crude Oil Markets (cont.)

- Export demand (calibrated CES demand)

$$X_{rr'} = \bar{x}_{rr'} \left(\frac{P_{r'}}{C_r(1+\tau_{rr'})} \right)^{\sigma_{r'}}$$

$\bar{x}_{rr'}$ Benchmark export from r to r'
$\sigma_{r'}$ Elasticity of substitution
$\tau_{rr'}$ Ad-valorem tariff on exports from r to r'.

- Product prices (CES price index)

$$P_r = \left(\sum_{r'} \theta_{r'r} \left(C_{r'}(1+\tau_{r'r}) \right)^{1-\sigma_r} \right)^{1/(1-\sigma_r}$$

$\bar{\theta}_{rr'}$ Benchmark value share of exports from r in r'

Optimal Sanctions

$$\min_{\tau_{\hat{r},r} \forall r \in \tilde{R}} C_{\hat{r}} Y_{\hat{r}}$$

subject to:

$$C, Y \in \mathscr{E}(\tau)$$

where $\hat{r}$ is the sanctioned region and $\tilde{R}$ are the set of regions in the sanctioning coalition.

Outline

- Mathiesen's modeling framework
- MPSGE
- CES functions and application
- Simulation analysis with GAMS and EXCEL
- WiNDC origins and motivation
- Equations of the GTAP-WiNDC model
- Policy application with GTAP-WiNDC: tariff quotas

Professor James Alm (1992):

The most interesting answer to any question which arises in economics is. . .

Professor James Alm (1992):

The most interesting answer to any question which arises in economics is. . .

"It depends."

Programming Constructs for Sensitivity Analysis

- *Environment variables* in GAMS provide string substitutions.

```
parameters      epsilon(r)      Price elasticity of demand for refined oil prod
                eta(r)          Price elasticity of supply for crude oil
                sigma(r)        Elasticity of substitution across crude oil var

epsilon(r) = %epsilon%;
eta(r) = %eta%;
sigma(r) = %sigma%;
```

- Default values are easily assigned.

```
$if not set epsilon    $set epsilon 0.2
$if not set eta        $set eta 0.5
$if not set sigma      $set sigma 4
```

- User-defined environment variables can be specified on the command line and solution values can be saved in GDX format.
- GDX output files can be merged and written to Excel worksheets or CSV files.

Outline

- Mathiesen's modeling framework
- MPSGE
- CES functions and application
- Simulation analysis with GAMS and EXCEL
- WiNDC origins and motivation
- Equations of the GTAP-WiNDC model
- Policy application with GTAP-WiNDC: tariff quotas

Collaboration with EDF (2017)

An important byproduct of our project will be an open-source dataset suitable for analysis of energy-economy-environment issues in North America. We begin with the national input-output table and downscale to the county level using regional economic statistics from the Bureau of Economic Analysis (sectoral value added and price household expenditure). We also employ data from Census Bureau's (foreign trade statistics) and International Trade Administration for bilateral trade statistics. Input-output tables will further be complemented by physical energy quantities and energy prices from the Department of Energy's State Energy Data System (SEDS) of EIA.

Overview

Research question: Can we build a set of transparent tools for producing subnational economic accounts for general equilibrium and input output analysis for the United States?

- Equilibrium analysis relies on *constructed* datasets.
- In this work, we reconcile national and state level economic data to produce a *micro-consistent* state level dataset for the United States.
- Publicly available regionalized accounts are not freely available, limiting the scope of equilibrium analysis.

Objectives

The Wisconsin National Data Consortium is being created to facilitate the coordination and implementation of:

- Open source build stream (which can be modified by users to produce their own version of regional social accounting matrices).
- Value shares, tax and trade margins based on public data.
- Estimated elasticities based on proprietary Census data with public code but restricted data.
- Connections to other international data sets.
- Accessible build stream which runs on NEOS (optimization server not requiring GAMS license).
- Clean connection to canonical models which run in both GAMS and Julia/JUMP.

The aim is to provide *options* for building a policy specific dataset and a foundational structure from which to base an analysis from.

Motivation

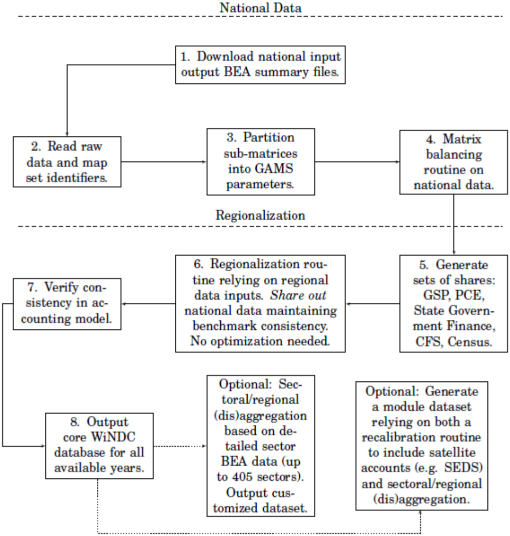
Existing subnational models have largely relied on a commercial database (IMPLAN) to characterize base year state and county-level economic activity in the United States.

- IMPLAN sells both state- and county-level national datasets which are based on public data
- Lack of transparency in regionalizing data. Outside options are expensive and proprietary. No way to look "under the hood".
- No mechanisms for understanding how data related assumptions impact model results.

The open-source tools for combining data and building a benchmark equilibrium database will be useful to many research groups across the country. Provide means for making more quantitative evidence based research possible.

Build Stream

Figure 2: Build Stream Process



National Tables

National level summary files from 1997-2017:

- Supply tables – byproduct matrices with aggregate imports and trade/transport margins.
- Use tables – includes aggregate intermediate inputs, total taxes, exports, and demand accounts (aggregate household, government purchases and investment).

Use of GAMS to define submatrices and partition into CGE based parameters.

Sector Disaggregation

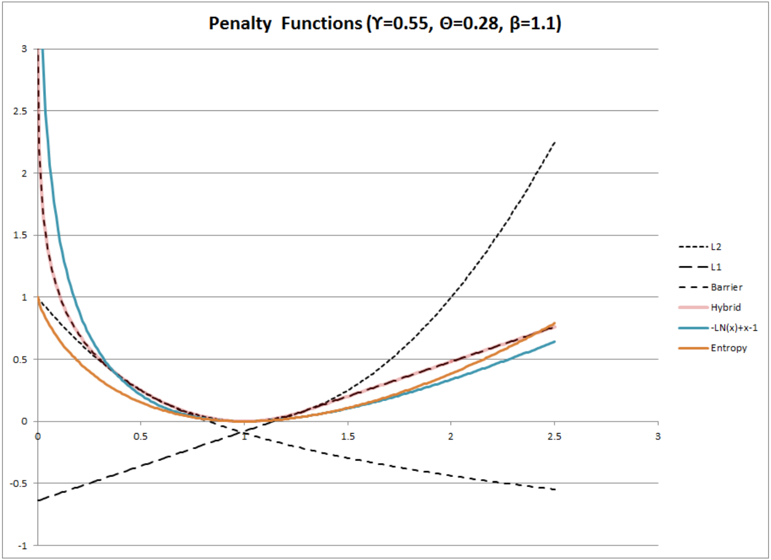
The routine provides options on the preferred level of sector disaggregation. Sector level detail is leveraged from the 2007 tables with 389 sectors. Level of disaggregation would depend on analysis. Options in the code include:

- *tot*: full disaggregation,
- *non*: no disaggregation,
- *eng*: energy related sectors,
- *agr*: agricultural sectors,
- *gtp*: GTAP disaggregation

For data in the 2007 tables, disaggregation shares are generated through linking disaggregate sector data with aggregate sector data through particular parameters. Data not in the disaggregate data (margins) are shared according to equal weight. Can use satellite data as well (oil and gas extraction).

- Huber's approach to matrix balancing incorporates a barrier function to assure that nonzeros in the source data remain nonzero in the estimated matrix. In the hybrid barrier method we retain Huber's loss function for increases from the target value and we add a log term to penalize values which go to zero:

$$\delta(a) = \begin{cases} 2\bar{a}\theta\left(\frac{a}{\bar{a}} - 1\right) & \text{for } \left(\frac{a}{\bar{a}} - 1\right) \leq \theta \\ \bar{a}\left(\frac{a}{\bar{a}} - 1\right)^2 & \text{for } -\gamma \leq \frac{a}{\bar{a}} - 1 \leq \theta \\ 2\bar{a}\gamma(1-\gamma)\ln\left(\frac{a}{\bar{a}}\right) & \text{for } \left(1 - \frac{a}{\bar{a}}\right) \geq \gamma \end{cases}$$


Penalty Functions (ϒ=0.55, Θ=0.28, β=1.1)
3
2.5
2
1.5
1
0.5
0
-0.5
-1
0
0.5
1
1.5
2
2.5
3
L2
L1
Barrier
Hybrid
-LN(x)+x-1
Entropy

Regionalization Process

The process to go from consistent national tables to state level tables relies on *sharing* data parameters. Shares are based on:

- gross state product (GSP)
- personal consumer expenditures (PCE)
- state government finance tables (SGF)
- USA trade statistics from Census
- commodity flow survey (CFS)

In the first four cases, data are given in aggregate categories. Categories are mapped to sectors in national data. Shares are generated such that:

$$\sum_r \delta_{yr,r,s} = 1 \quad \forall \quad (yr, s)$$

Regionalization Process

- Use GSP shares to separate production data: sectoral supply with byproducts, intermediate demand and value added. Split aggregate value added based on labor and capital accounts in GSP data.
- Use PCE shares to separate household final consumption.
- Use SGF shares to separate government expenditures.
- GSP shares separate investment demand.
- USA trade shares based on Census data to separate state exports.
- For a given year then, total domestic absorption must equal:

$$= HHDem_{r,g} + GovDem_{r,g} + Inv_{r,g} + \sum_{s} IDem_{r,g,s}$$

- Generate implicit shares based on absorption totals to enforce identities:

$$= Abs_{r,g} / \sum_{rr} Abs_{rr,g}$$

- Use implicit shares to separate imports and margin demand.

Regionalization Process

In order to maintain zero profit and market clearance in the data, we determine demand/supply from/to the state vs. national markets by imposing *regional purchase coefficients* based on commodity flow survey data or a gravity model of trade.

- Regional purchase coefficients (RPC) are found by assigning aggregate categories in CFS data to blueNOTE sectors or through estimated bilateral trade flows. The dataset provides a metric on how much of a given good is retained in a given state or shipped to other states.
- $RPC_{r,g} \in [0, 1]$. I.e. an $RPC_{r,g} = 0.4$ would indicate 40% of a given good's domestic demand was sourced in the state. The rest came from the national market.

State level or national level domestic demand is defined by either the supply or demand side of the market to maintain zero profit in either the export or absorption markets.

Margins are supplied by both the state and national markets.

Trade

The benchmark dataset is structured for either a pooled national market or gravity based estimates. Explicit bilateral trade flows cannot be determined using CFS data:

- Wittwer (2017) shows that CFS data provide information on the value of goods between transport nodes, which may or may not be in line with production origins or consumption destinations.
- Points to need of gravity based estimates.

Bilateral Trade

Estimated with Canadian D-Level input output data for 2014 for each blueNOTE sector. Trade from region i to j depends on economic forces in both origin and destination nodes, and forces that aid or restrict the flow of goods from origin to destination.

$$lnY_{ij} = \beta_0 + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) + \beta_3 \ln(Dist_{ij}) + \sum_f \beta_f X_{ij}^f + \epsilon_{ij}$$

	(1) *all*	(2) *adm*	(3) *agr*	(4) *air*	(5) *alt*	(6) *amb*	(7) *amd*
lnFromGDP	0.713***	1.014***	0.437***	1.132***	0.428***	0.597***	0.703***
	(0.0638)	(0.0782)	(0.0687)	(0.0617)	(0.0637)	(0.0852)	(0.0540)
lnToGDP	0.549***	0.810***	0.551***	0.739***	0.369***	0.553***	0.768***
	(0.0603)	(0.0816)	(0.0735)	(0.0813)	(0.0604)	(0.0795)	(0.0637)
lnDist	-1.339***	-1.370***	-1.322***	-1.089***	-0.759***	-1.469***	-1.220***
	(0.121)	(0.145)	(0.148)	(0.141)	(0.128)	(0.167)	(0.0990)
Contiguity	-0.302	0.284	0.188	-0.575	-0.185	-0.532	0.0805
	(0.272)	(0.393)	(0.302)	(0.349)	(0.286)	(0.412)	(0.254)
Language	0.483***	0.393*	0.798***	0.719***	0.199	0.625***	0.674***
	(0.159)	(0.224)	(0.196)	(0.214)	(0.147)	(0.225)	(0.161)
Constant	-6.618***	-8.854***	-5.622***	-11.26***	-8.374***	-5.381***	-7.816***
	(0.909)	(1.423)	(0.909)	(1.436)	(0.825)	(1.268)	(0.781)
Observations	49100	700	2000	300	600	400	200
R2	0.439	0.594	0.510	0.522	0.415	0.552	0.585

Notes: Standard errors, clustered by origin destination pairs, are in parentheses with * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Sectors are described as follows: all - all sectors, adm - administrative and support services, agr - farms, air - air transportation, alt - apparel and leather and allied products, amb - ambulatory health care services, amd - accommodation.

Data Overview

The build routine provides:

- Social accounting matrices for all 50 states plus D.C. from 1997-2017.
- Based on summary files of 71 sectors.
- Option for disaggregation using the 2007 389 sectoring scheme and additional satellite accounts.
- Regionalization achieved mainly regional level gross state product and expenditure accounts.
- Trade is imposed in national pooled market using regional purchase coefficients generated by commodity flow survey data or through gravity based estimates.
- Option for recalibrating dataset to match totals from satellite accounts.

Table 1: WiNDC Data Sources

Source	Description	ID	URL	Years
Bureau of Economic Analysis	Supply and Use Tables	*BEA*	`https://www.bea.gov/industry/io_annual.htm`	1997-2017
	Gross State Product	*GSP*	`https://www.bea.gov/newsreleases/regional/gdp_state/qgsp_newsrelease.htm`	1997-2016
	Personal Cosumer Expenditures	*PCE*	`https://www.bea.gov/newsreleases/regional/pce/pce_newsrelease.htm`	1997-2017
Census Bureau	Commodity Flow Survey	*CFS*	`https://www.census.gov/econ/cfs/`	2012
	State Government Finance	*SGF*	`https://www.census.gov/programs-surveys/state/data/tables.All.html`	1997-2016
	State Exports/Imports	*UTD*	`https://usatrade.census.gov`	2002-2016
Energy Information Administration	State Energy Data System	*SEDS*	`https://www.eia.gov/state/seds/`	1963-2016

Notes: Years indicates the *usable* years of available data across data sources. For instance, state level gross product is available for 2017 but has many hidden entries.

First version of WiNDC featured a state level dataset with a single representative agent by region.

- Provided means for spatially denominated distributional analysis, but not within consumer types.
- A key advantage of IMPLAN was its disaggregation of regional consumer demands and incomes by household income groups.
- Many ways to go about this type of disaggregation. Incomes vs. expenditures.
- We approach this problem from the income side. Key challenges: denominate reasonable transfer income, understand income tax liabilities, savings, capital ownership vs. demands, salaries and wages.
- Additional wrinkle: static vs. steady state calibration.
- Income elasticities used to separate household level commodity expenditures.

Household recalibration routine:

- Two versions of a household dataset is produced. One primarily based on the Current Population Survey (CPS), and the other based primarily on IRS's Statistics of Income (SOI).
- Both versions use a bit of information from the other. Transfers and capital gains.
- Roughly comparable with 5 household types by region. Households vs. returns.

Income Balance in the Benchmark Equilibrium

Original regional representation (subscripted by r) – limited by information in the reference input output tables:

$$cons_r + inv_r = wages_r + cap_r + other_r \quad \forall \quad r$$

- Investment based on location of state level investment demands. May not follow location of entity actually doing the investing.
- Wages and capital income based on sectors in a given state doing the demanding. Again, same issue. Furthermore, they are gross of taxes.
- Other is a closure parameter – all the stuff that can't be explained by consumption, investment, wages and capital.

Obvious issues when thinking about welfare impacts.

Toward a Better Income Balance Representation

While regional representation may limit ability to do reasonable welfare analysis, it does provide useful *control totals* that are consistent with both the National Income and Product Accounts (NIPA) and accounting identities for the rest of the economy.

This work seeks to reconcile the issues outlined in the previous slide. Move toward the following income balance representation:

$$cons_{rh} + tax_{rh} + save_{rh} = wages_{rh} + cap_{rh} + trn_{rh} \quad \forall \quad (r, h)$$

- Break out each category by region and household income type (h).
- Estimate savings for household-region pairs investing in new capital.
- Estimate wage and extant capital endowments consistent with where people actual live and work. Incorporate income taxes into WiNDC structure.
- Break out the "other" category into cash payment transfers consistent with benefits programs in US. Assume all transfers are between households and government. No intra-household transfers assumed here.

Datasets Used in Disaggregation

Current Population Survey (CPS)	Statistics of Income (SOI)
Survey dataset	Administrative dataset
Observation level micro-use dataset	Only have access to precompiled state-level statistics -- micro-use public file prohibitively expensive, no spatial identifier
Census defined households	Number of tax returns
Respondents report taxable and non-taxable income	Only reports taxable income (including transfers)
No income tax payments	All income tax payments for those paying taxes
Denominates all major cash trasfer payments from government programs	Transfer payments that are taxed (very limited)
Wages and salaries	Wages and salaries
Interest income (no capital gains)	Interest income (with capital gains)
Savings limited to retirement income	Savings limited to retirement income

CPS Categories (2016)

Income Balance	CGE category	Data category	Total/%
Income	Labor		$7,883.57
		Wages and salaries	100%
	Capital		$952.66
		Self-employment (farm and nonfarm)	43.54%
		Interest	31.26%
		Dividends	13.58%
		Rents	10.37%
		Other income	1.24%
	Transfers		$1,113.66
		Social security	69.03%
		Educational assistance	6.09%
		Veteran's benefits	5.21%
		Supplemental security	4.44%
		Survivor's income	4.05%
		Disability	3.36%
		Financial assistance	2.43%
		Child support	2.09%
		Unemployment compensation	1.65%
		Workers compensation	1.05%
		Public assistance or welfare	0.61%
Expenditures	Savings		$545.81
		Retirement income	100%

SOI Categories (2016)

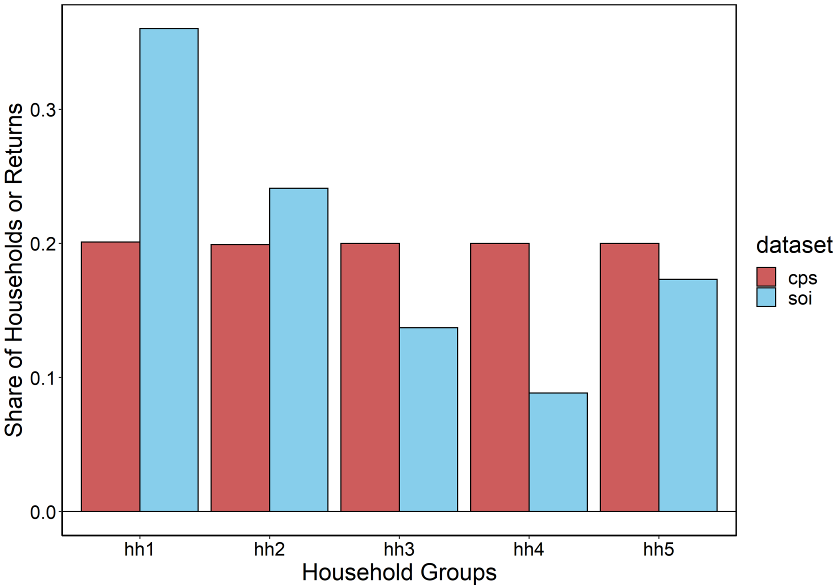
Income Balance	CGE category	Data category	Total/%
Income	Labor		$7,114.21
		Wages and salaries	100%
	Capital		$3,073.52
		Rental real estate, royalties, partnerships, S corporations, trusts, etc.	21.69%
		Capital gains	19.41%
		Business income	11.22%
		Ordinary dividends	7.89%
		Qualified dividends	6.27%
		Taxable interest	2.91%
	Transfers		$310.56
		Social security benefits (taxable portion)	91.61%
		Unemployment compensation	8.39%
Expenditures	Taxes		$1,552.37
		Federal	97.86%
		State and local tax credits	2.14%
	Savings		$37.33
		Self-employed SEP, SIMPLE, and qualified plans	65.02%
		IRA deduction	34.98%

Households vs. Number of Returns

In what follows, we aggregate households into 5 groups that *roughly* correspond with one another. Comparison between two datasets not perfect.

ID	CPS	SOI
hh1	under $24,000	under $25,000
hh2	$24,000 to $45,442	$25,000 to $50,000
hh3	$45,442 to $74,567	$50,000 to $75,000
hh4	$74,567 to $120,951	$75,000 to $100,000
hh5	over $120,951	over $100,000

Households vs. Number of Returns



Structure of Recalibration Routine

4 step process:

1. Solve for steady state equilibrium investment demands (if option is selected – static vs. steadystate).
    - Important because investment levels tie directly to the income balance constraint for households in the form of savings. Considering this upfront circumvents issues down the line.
2. Solve income routine for aggregated regions (here Census regions).
3. Solve income routine at the state level enforcing control totals at the aggregated region level.
4. Solve expenditure routine at the state level.

Successive calibration is akin to the LES calibration used in SAGE. Enhances reliability in when solving a larger model.

GTAP Integration

- Sectoral reconciliation: 33 sectors representing aggregation of both GTAP and WiNDC structure
- 50 states plus DC
- Any number of GTAP regions (free version with 11 regions)
- Aggregate trade aligns with GTAP. Use matrix, final demand and household endowment structgure aligns with WiNDC.

Sectoral Mapping by Alla Golub for GTAPWiNDC

U.S. economy is represented with 71 BEA sectors in the WiNDC data system mapped to North American Industry Classification System (NAICS), while economic activities are grouped into 65 sectors in the GTAP database and mapped to Central Product Classification (CPC) or International Standard Industrial Classification (ISIC). To embed a subnational model for the U.S within the GTAP framework, a common set of economic sectors is needed. Careful assessment of these classificaitons produces a 33 sector mapping for GTAP-WiNDC project.

Outline

- Mathiesen's modeling framework
- MPSGE
- CES functions and application
- WiNDC origins and motivation
- Equations of the GTAP-WiNDC model
- Policy application with GTAP-WiNDC: tariff quotas

The GTAP-WiNDC Model

```
$title  The GTAPWiNDC Model

*       Use the 2014 dataset with households defined by CPS:

$if not set ds $set ds gtap_cps_2014

$include gtapwindc_data

$ontext
$model:gtapwindc

$sectors:
        Y(g,r,s)        ! Production (includes I and G)
        C(r,s,h)        ! Consumption
        X(i,r,s)        ! Disposition
        Z(i,r,s)        ! Armington demand
        FT(f,r,s)       ! Specific factor transformation
        M(i,r)          ! Import
        YT(j)           ! Transport
```

The GTAP-WiNDC Model (cont.)

```
$commodities:
        PY(g,r,s)       ! Output price
        PZ(i,r,s)       ! Armington composite price
        PD(i,r,s)       ! Local goods price
        PM(i,r)         ! Import price
        P(i,r)          ! National goods price
        PC(r,s,h)       ! Consumption price

        PF(f,r,s)       ! Primary factors rent
        PS(f,g,r,s)     ! Sector-specific primary factors
        PT(j)           ! Transportation services

$consumers:
        RH(r,s,h)       ! Representative household
        GOVT(r)         ! Public expenditure
        INV(r)          ! Investment
```

Central Variables: Activity levels

$Y_{g,r,s}$	Production (includes I and G)
$C_{r,s,h}$	Consumption
$X_{i,r,s}$	Disposition
$Z_{i,r,s}$	Armington demand
$FT_{f,r,s}$	Specific factor transformation
$M_{i,r}$	Import
YT_{j}	Transport

Central Variables: Prices

$PY_{g,r,s}$	Output price
$PZ_{i,r,s}$	Armington composite price
$PD_{i,r,s}$	Local goods price
$P_{i,r}$	National goods price
$PC_{r,s,h}$	Consumption price
$PF_{f,r,s}$	Primary factors rent
$PS_{f,g,r,s}$	Sector-specific primary factors
$PM_{i,r}$	Import price
PT_{j}	Transportation services

Central Variables: Income levels

$RH_{r,s,h}$	Representative household
$GOVT_r$	Public expenditure
INV_r	Investment

Market Clearance Conditions

- Output price – $PY_i \quad i = \{g, r, s\}$

$$\sum_{s=\{g,r,s\}} Y_s \underbrace{a^{PY,Y}_{i\times s}}_{\texttt{o:vom}} + \sum_{s=\{i,r,s\}} X_s \underbrace{a^{PY,X}_{i\times s}}_{\texttt{i:vom}} \geq \sum_{c=\{r\}} \underbrace{D^{PY,GOVT}_{i\times c}}_{\texttt{d:vom}} + \sum_{c=\{r\}} \underbrace{D^{PY,INV}_{i\times c}}_{\texttt{d:vom}}$$

- Armington composite price – $PZ_i \quad i = \{i, r, s\}$

$$\sum_{s=\{g,r,s\}} Y_s \underbrace{a^{PZ,Y}_{i\times s}}_{\texttt{i:vafm}} + \sum_{s=\{r,s,h\}} C_s \underbrace{a^{PZ,C}_{i\times s}}_{\texttt{i:cd0}} + \sum_{s=\{i,r,s\}} Z_s \underbrace{a^{PZ,Z}_{i\times s}}_{\texttt{o:a0}} \geq 0$$

- Local goods price – $PD_i \quad i = \{i, r, s\}$

$$\sum_{s=\{i,r,s\}} X_s \underbrace{a^{PD,X}_{i\times s}}_{\texttt{o:xd0}} + \sum_{s=\{i,r,s\}} Z_s \underbrace{a^{PD,Z}_{i\times s}}_{\texttt{i:xd0}} \geq 0$$

- National goods price – $P_i \quad i = \{i, r\}$

$$\sum_{s=\{i,r,s\}} X_s \underbrace{a^{P,X}_{i\times s}}_{\texttt{o:xn0}} + \sum_{s=\{i,r,s\}} Z_s \underbrace{a^{P,Z}_{i\times s}}_{\texttt{i:nd0}} + \sum_{s=\{i,r\}} M_s \underbrace{a^{P,M}_{i\times s}}_{\texttt{i:vxmd}} + \sum_{s=\{j\}} YT_s \underbrace{a^{P,YT}_{i\times s}}_{\texttt{i:vst}} \geq 0$$