The GTAP-WiNDC Framework

Thomas F. Rutherford

University of Wisconsin – Madison

May 19, 2022

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.

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.

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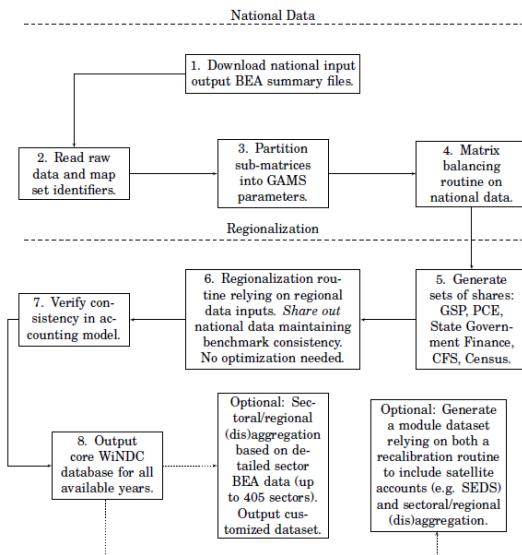
Existing subnational models have largely relied on a commercial database (IMPLAN) to characterize base year state and county-level economic activity in the United States.

- 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.

Figure 2: Build Stream Process

National Data

1. Download national input output BEA summary files.

2. Read raw data and map set identifiers.

3. Partition sub-matrices into GAMS parameters.

4. Matrix balancing routine on national data.

Regionalization

7. Verify consistency in accounting model.

6. Regionalization routine relying on regional data inputs. *Share out national data maintaining benchmark consistency. No optimization needed.*

5. Generate sets of shares: GSP, PCE, State Government Finance, CFS, Census.

8. Output core WiNDC database for all available years.

Optional: Sectoral/regional (dis)aggregation based on detailed sector BEA data (up to 405 sectors). Output customized dataset.

Optional: Generate a module dataset relying on both a recalibration routine to include satellite accounts (e.g. SEDS) and sectoral/regional (dis)aggregation.

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.

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).
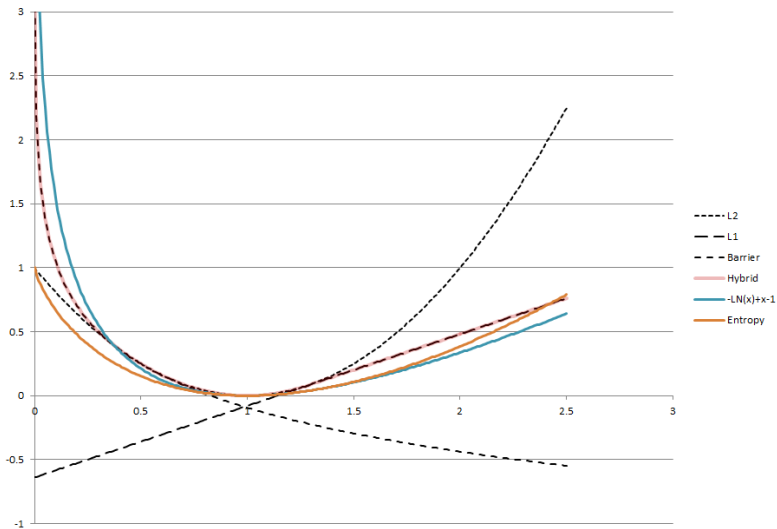
Huber's approach to matrix balancing incorporates a barrier function to assure that nonzeros in the source data remain nonzero in the estimated matrix.

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:

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

Penalty Functions (ϒ=0.55, Θ=0.28, β=1.1)

Legend:
- L2
- L1
- Barrier
- Hybrid
- -LN(x)+x-1
- Entropy

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:

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

- 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.

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.

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.

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.

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.

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

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%

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%

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
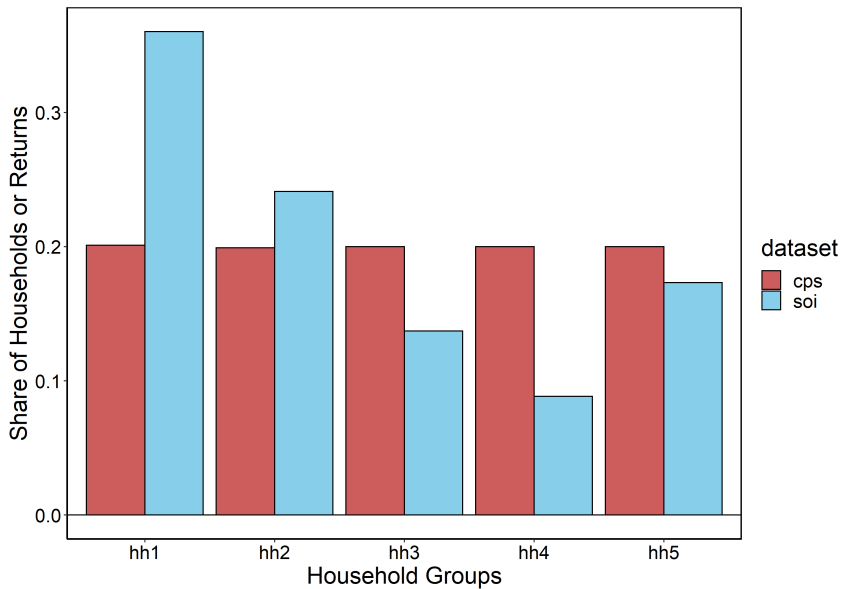
4 step process:

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.

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.

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.

```
$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
```

```
$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
```

Representation of tariff quota requires some additional modeling: in-quota and out-of-quota trade flows are distinguished:

```
set     q         Quota levels / in        "In-quota trade",
                                out        "Out-quota trade" /,

$sectors:
...
        B(i,rr,r,q)        ! Bilateral export (for quota constrained trade)


$commodities:
...
        PB(i,rr,r)         ! Bilateral export price
        PQ(i,rr,r,q)       ! Quota rent on bilateral trade


$prod:B(i,rr,r,q)  s:0
        o:PB(i,rr,r)       q:vmcif(i,rr,r)

        i:P(i,rr)          q:vxmd(i,rr,r)  p:pvxmd(i,rr,r)
+               a:GOVT(rr)  t:(-rtxs(i,rr,r))
+               a:GOVT(r)   t:(rtms_q(i,rr,r,q)*(1-rtxs(i,rr,r)))

        i:PT(j)            q:vtwr(j,i,rr,r)  p:pvtwr(i,rr,r)
+               a:GOVT(r)   t:rtms_q(i,rr,r,q)

        i:PQ(i,rr,r,q)$quota(i,rr,r,q)  q:vxmd(i,rr,r)
```
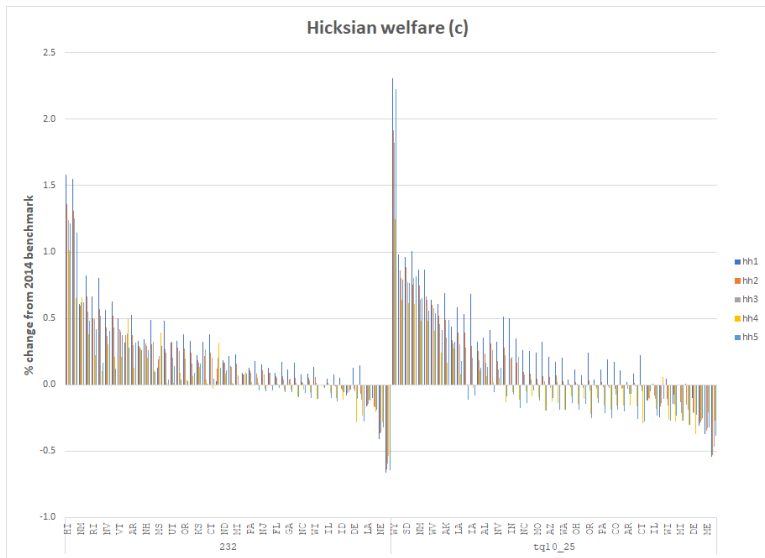
Steel Tariff Incidence by State and Income Quintile

Hicksian welfare (c)

← → C | ⓘ Not secure | www.windc.wisc.edu

UNIVERSITY OF WISCONSIN-MADISON

**WISCONSIN NATIONAL DATA
CONSORTIUM**
WiNDC: Open-Source Datasets for Economic Analysis

HOME   GETTING STARTED   CONSORTIUM   PEOPLE   RESOURCE CENTER ⌄   DOWNLOADS   CONTACT US

—
ABOUT WINDC

The Wisconsin National Data Consortium (WiNDC) facilitates the creation and documentation of open source multisectoral economic datasets for US states and counties. These datasets are created by open source computer programs which can be run on NEOS. The datasets are provided together with canonical general equilibrium models which can provide the starting point for quantitative policy analysis.

The current open-source dataset is a byproduct of a recently completed research project conducted by Thomas F. Rutherford and Andrew Schreiber (both from the University of Wisconsin-Madison) with Gökçe Akin-Olçum (from the Environmental Defense Fund) and Christoph Böhringer (from the University of Oldenburg). It is suitable for analyzing economy-wide issues in North America.

WiNDC was organized and is hosted by the Department of Agricultural & Applied Economics at the University of Wisconsin-Madison.