Generating Spatial Time Series on Interstate Inter-industry Freight Flows

Kieran Donaghy
Jialie Chen
Department of City and Regional Planning
Cornell University
2011.10.14

Acknowledgement

- The research reported in this paper has been supported by grant RD – 8342831 from the U.S. Environmental Protection Agency (EPA) STAR program, “Regional Infrastructure/Air Quality Planning and Global Change.”

Outline

- Motivation
- Introduction–Why three region (Midwest, Northeast and New England) model?
- Practical Problems
- General Procedures–Base year and coefficient updates (REIM)
- Model outcomes–Two states case
- Conclusion

Motivation

- We are interested in how commodity flow affects air quality, especially among the following questions:

1) How will emerging technologies and infrastructure changes (including transportation) and global climate and air pollution policies impact regional air quality?

Motivation Cont.,

2) How will urban and regional landscapes evolve in response to changes in production and distribution at regional and global scales? How will associated changes in land use affect emissions and air quality?

- *And forecasting freight movement is critical to answer those questions.*

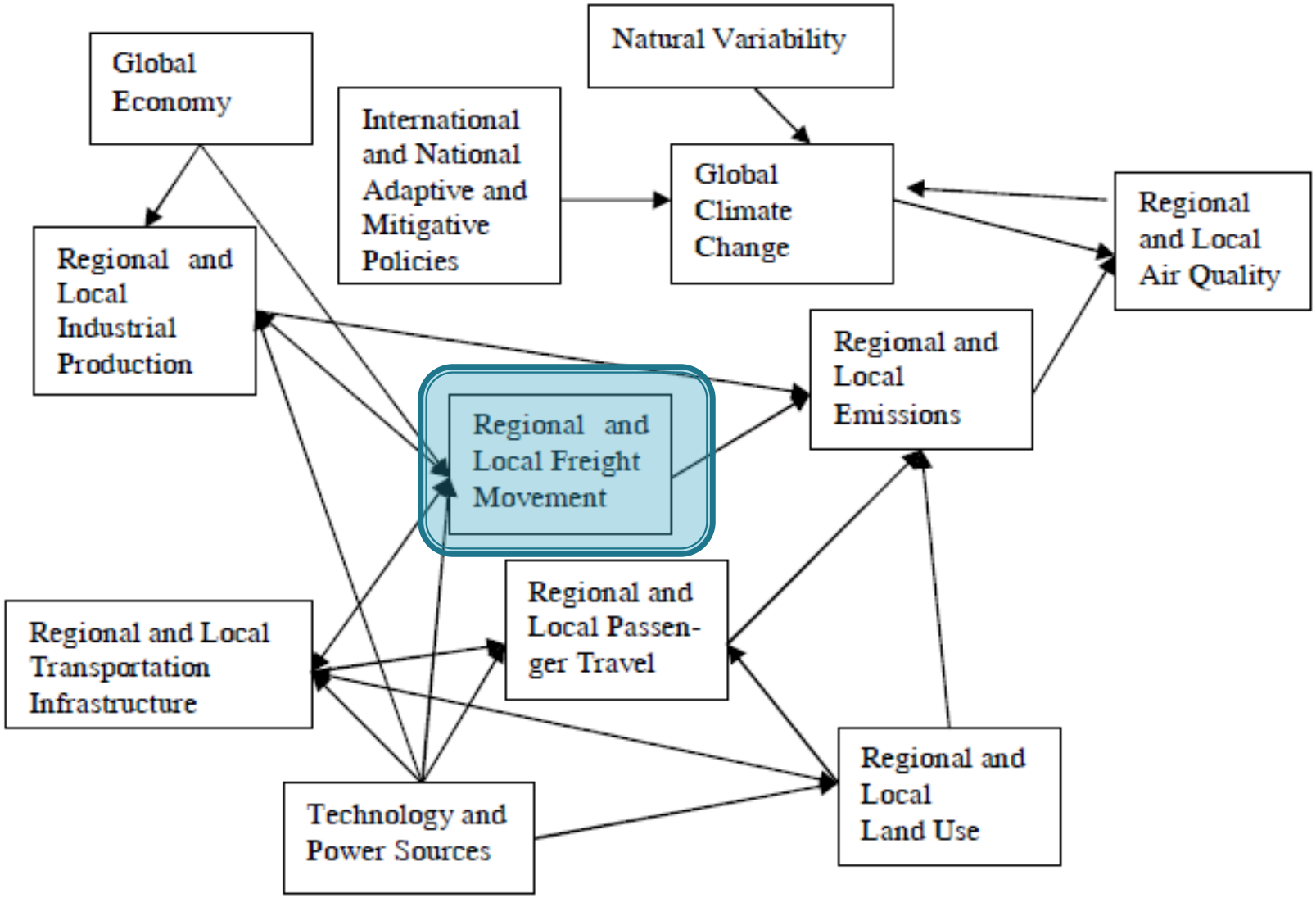
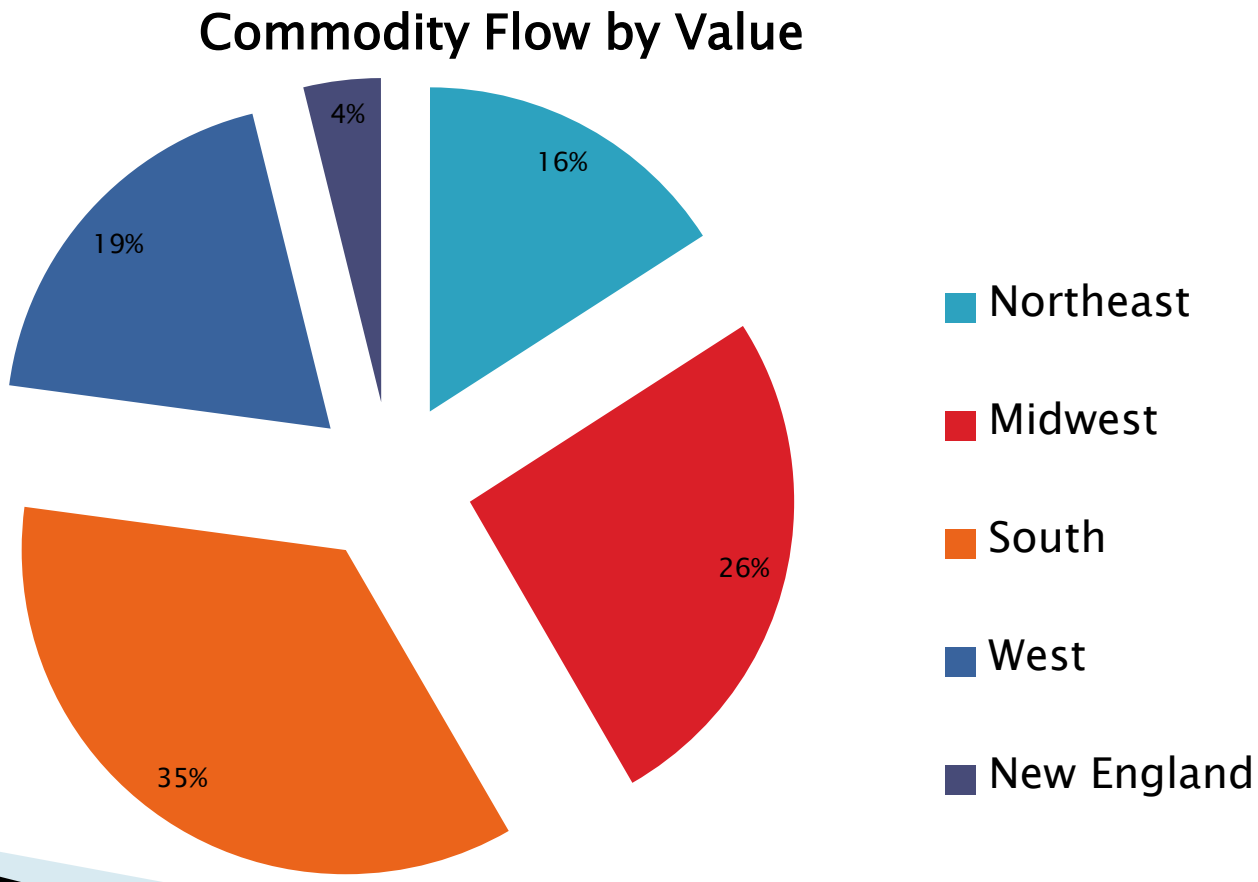


Figure 1. Causal Relations between Factors Affecting Regional and Local Air Quality

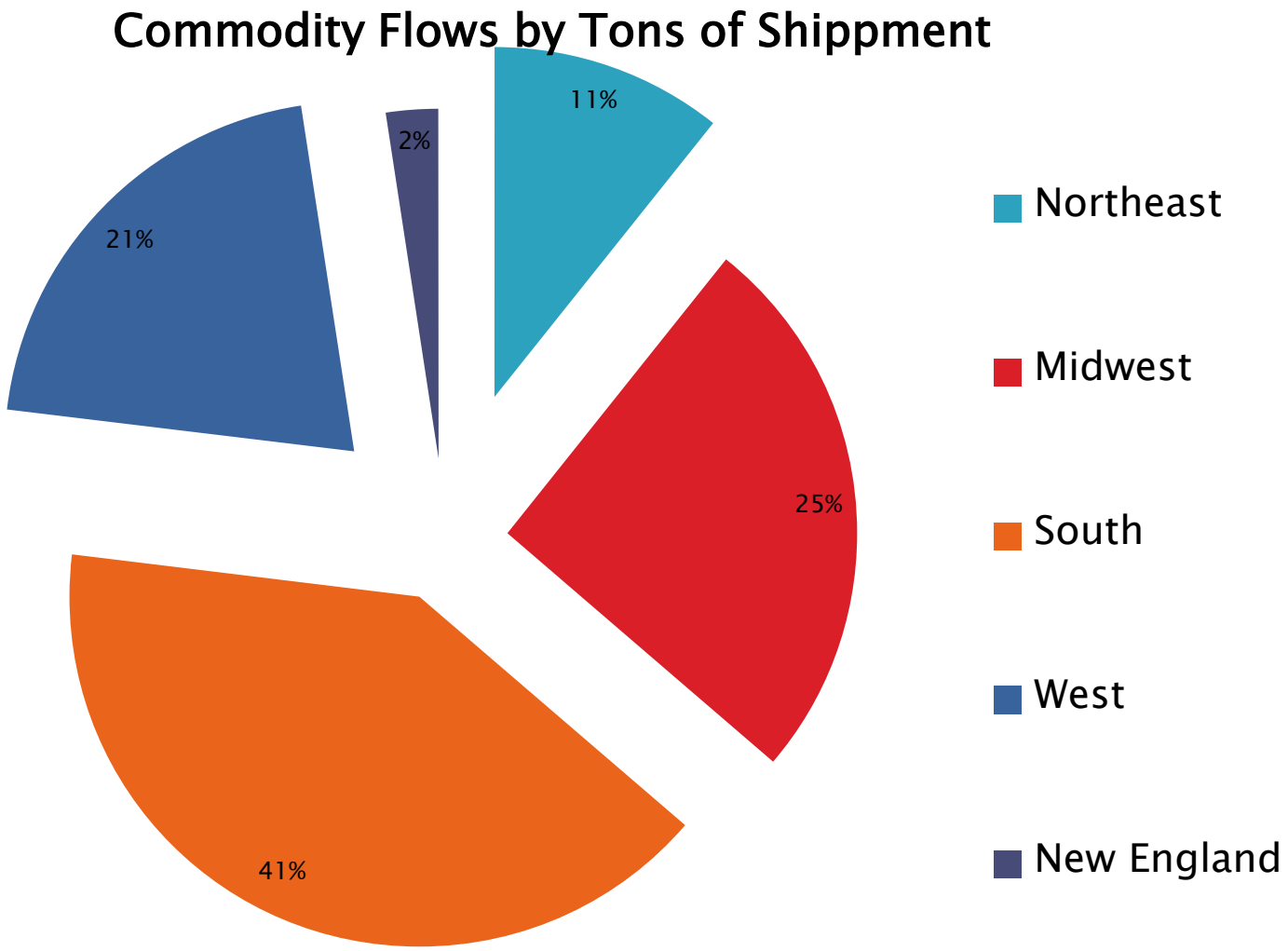
Motivation Cont.,

- Dynamic commodity flow model will be estimated to analyze the inter-regional freight movement
- We will focus on thirteen states from three regions of US, with thirteen industries in each state.

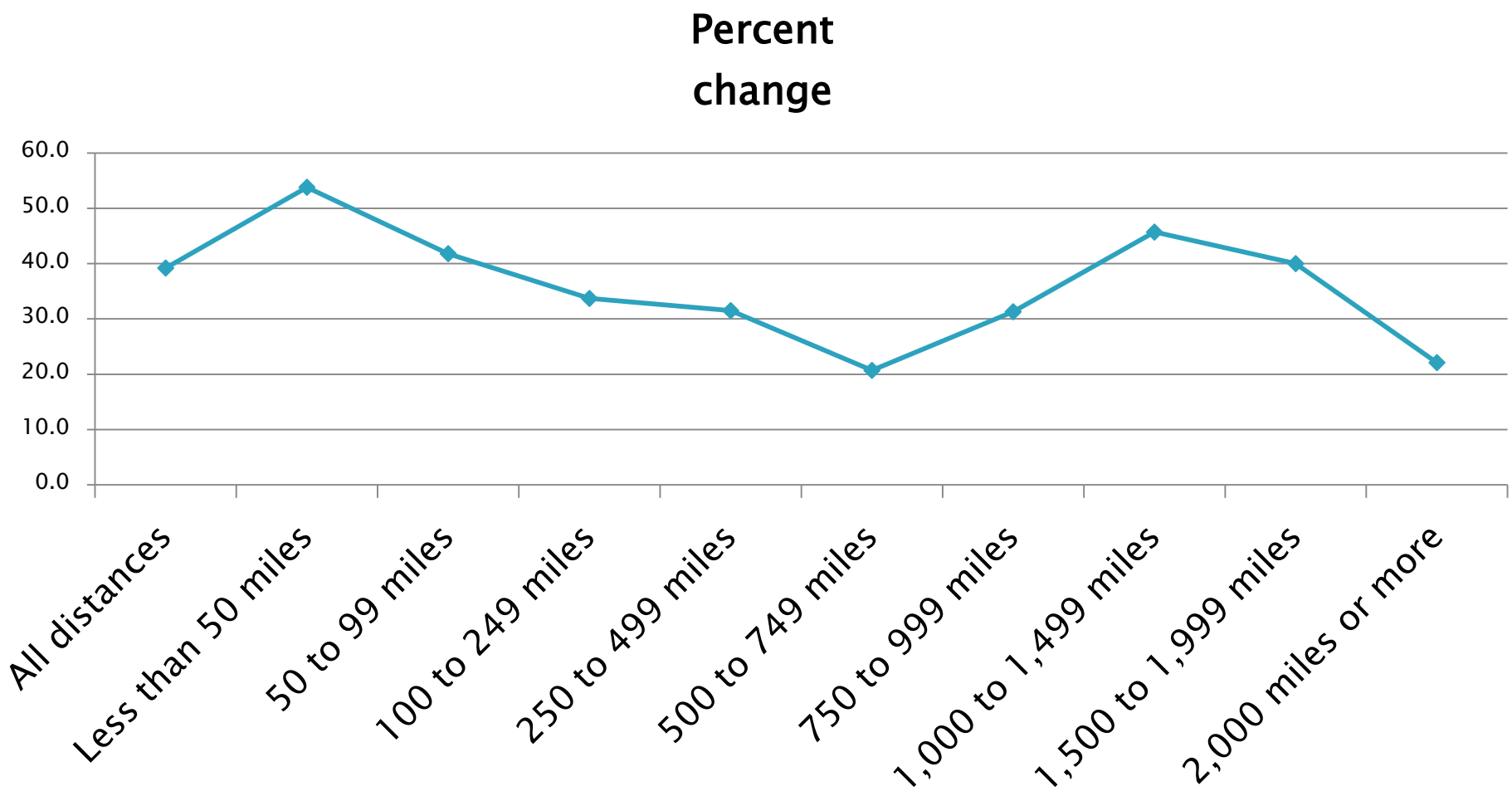
Introduction–Why three region model?



Introduction Cont.,



Overall Commodity Flow Data



Source: Bureau of Transportation Statistics

Introduction Cont.,

- Summarization

1) Three regions (Mid-West, Northeast and New England) occupy a considerable share of the total value for commodity flows
2) Inter-regional trade has become more and more active.
3) Dynamic flow model offers us a valuable venue to study inter-regional economics activities.

Practical Problems

- Lack of time series data for commodity flows to estimate dynamic commodity flow models.

- ✓ *A Regional Econometric Input-Output Model (REIM) could be constructed to generate the missing data.*

- ✓ *However, REIM requires reliable base year input-output information.*

Practical Problem Cont.,

- Lack of reliable base year data:
- State-to-state commodity flow estimates published by US Bureau of Transportation Statistics (BTS) are not reliable because :

1) High statistical variability and thus lack of confidence;
2) Reporting shipment origin and destination rather than manufacturing origin.

Estimation of Data from the Base Year

- *Randy Jackson et al's (2005) approaches will be applied to recover the base year data.*
- *An alternative approach is developed to generalize the distance volume relationships from BTS data in order to smooth out irregularities.*

Estimation of Data from the Base Year

- Generating the distribution of known regional domestic exports
- Predicted flow of commodity i from region m to n:

$$\hat{Y}_i^{mn} = \frac{(w_i^n)^{\varphi_i} \exp(-\delta_i d^{mn})}{\sum_n (w_i^n)^{\varphi_i} \exp(-\delta_i d^{mn})} y_i^{m\bullet}$$

w as the weight reflecting demand of imports from region n and d indicates the distance; y is observed total exports from BTS commodity-specific summary data.

- δ and ψ are estimated to minimizes the difference between predicted commodity flow and the actual commodity flow.

Commodity-by-Industry Accounting

	Commodity	Industry	Final Demand	Total
Commodity		U	E=F+X(M)	Q
Industry	V			X
Value Added		W		
Total	Q'	X'		

Commodity-by-Industry Accounting

- Use Matrix (U): Column industry purchase of row commodity
- Make Matrix (V): Column commodity output of row industry
- Value Added (M): Payment sectors
- Final Demand (E): Consumption, Investment, Government Spending and Net Export.

Commodity-by-Industry Accounting

- Traditional I-O definition
- $X_i = V_{i1} + V_{i2} + \ldots + V_{im}$ →Total industry outout
- $Q_i = U_{i1} + U_{i2} + \ldots + U_{im} + E_i$ →Total commodity output
- Define $B = U(X\hat{})^{-1}$, then we have $Q = BX + E$
- Commodity based approaches
- To determine the proportion of output of industry i contributed to commodity j, written as:
- $C = V'(X\hat{})^{-1}$ →Industry output proportion

Base Year I-O Coefficient Matrix

- $X\hat{} = C^{-1}V'$
- $X = X\hat{}\,i = C^{-1}V'\,i$ *i as unit vector*
- $Q = BX + E \rightarrow Q = BC^{-1}Q + E$
- Compare that with traditional I-O approaches:
- $X = AX + Y$
- $Ac = BC^{-1} = U(V')^{-1}$

Base Year I-O Coefficient Matrix

- From Miller and Blair (1985), I-O coefficient matrix for base year could be represented as :

$$Ac = U(V')^{-1} ,$$

under Commodity-Based Technology Assumption.

Updating I-O Coefficient Matrix

With base year coefficients in hand, we could proceed to the estimation of Regional Econometrics Input-Output Model (REIM) and thus have coefficients updated on a yearly basis.

- Data Sources for REIM
- REIM formulation
- Solving the model

Data Sources for REIM

- Final Demand
  - Gross Regional Output
  - ✓ *Available from Bureau of Economic Analysis*
  - Consumption
  - Investment
  - Government Spending
  - ✓ *Available from Bureau of Economic Analysis*
  - Net Export

Consumption

- No local data available
- Four regressions model will be constructed on a national level with the following break down:

autos and auto parts	non-auto durable goods	nondurable goods	services

Consumption Cont.,

- Estimations will be made by substituting regional data into the models

Model For Non-durable Goods	Dependent Variable	log(Consumption3/Population)
	Coefficient	t-Value
Constant	-15.96	-12.81
log(Income/Population)	1.41	4.17
log(CPI-Apparel/CPI)	0.94	1.33
log(Unemployment)	-0.62	-2.07
R^2	0.86	

Investment

- Investment could be classified into: residential, nonresidential and equipment.
- Data are usually available on national level from various sources, including US Census Bureau.
- Regional data could be estimated by prorating with number of construction permits and employment accordingly.

Net Export

- No direct data available on regional level.
- Could be estimated with the following equation:

$$Y=C+I+G+X$$

Sectoral Data

- Income
- Employment
- ✓ *Available from the Bureau of Labor Statistics*
- Output
- ✓ *Estimated as the total value of sales to match input-output table in the base year.*

Income

- Calculated with the following relationship:

Personal Income = Total Income + Property Income + Transfer Payments – Contributions to Social Insurance + Residential Adjustment to Income

Output

- Manufacturing Sectors

Estimated from value added, costs of materials and payroll series from Census of Manufacturers, the Annual Survey of Manufacturers and Bureau of Economic Analysis.

- Non-Manufacturing Sectors

Prorate national output series available from the Bureau of Labor Statistics by employment with the assumption that local output per worker is equivalent to the national level.

REIM Formulation

- Three-Region Model Specification:

Project Horizon	1977-2007 (31 years)
Model Sizes	169 sets of linear equation system
Industry Detail	13 industries with projections of output and inter-industry/regional commodity flow
Selected Exogenous Variables	Gross Regional Products
	Personal consumptions
	Residential and non-residential investment
	Federal government expenditures
	State and local government expenditures
	Population
	Unemployment rate
	Personal income
	Consumer Price Index (CPI)

Model Specification Cont.,

- Sector Definition

Sector	Description	NAICS Code
1	Agriculture, forestry, fishing, and hunting	11
2	Mining	21
3	Construction	23
4	Food product manufacturing	311
5	Chemical manufacturing	325
6	Primary metal manufacturing	331
7	Fabricated metal product manufacturing	332
8	Machinery manufacturing	333
9	Computer and electronic product manufacturing	334,335
10	Transportation Equipment	336
11	Other Non-durable Manufacturing	312-316,322-324,326
12	Other Durable Manufacturing	321,327,337,339
13	TCU, Service and Government Enterprises	42,44,45,48,49,51-56,61,62,71,72,81

REIM Formulation

- Relationship of estimated output (Z) and actual output (X)

$$Z_{REIM} = AX + Y$$

- X and Z are generally not identical unless for the base year.
- The systematic variation in the difference between Z and X can be accounted for by the systematic variation in a set of exogenous variables.

REIM Formulation Cont.,

- Stochastic Output Adjustment

$$\log(X_{i,t}/Z_{i,t})=\alpha_0+\alpha_z(Z_{i,t-1}/X_{i,t-1})+\alpha_g G_{it}+\xi_{it}$$

- ✓ *G is the time series of exogenous variables, generally including consumption in service, investment in equipment, government spending and net export.*

$$\beta_{it}\equiv X_{i,t}/Z_{i,t} =\exp(\alpha_0+\alpha_z(Z_{i,t-1}/X_{i,t-1})+\alpha_g G_{it}).$$

- $X_i=\beta_i \wedge Z_i$
- $X=\beta \wedge AX+ \beta \wedge Y$
- $X=[(I-\beta \wedge A)^{-1} \beta \wedge]Y$

REIM Formulation Cont.,

- Productivity Equation

$$\log(X_{i,t}/N_{i,t})=\alpha_0+\alpha_z(N_{i,t-1}/X_{i,t-1})+\alpha_g G_{it}+\xi_{it}$$

- ✓ *N is the total sectorial employment and G includes hours worked, unemployment and earnings etc.,*

$$\gamma_{it} = \exp(\alpha_0+\alpha_z(N_{i,t-1}/X_{i,t-1})+\alpha_g G_{it})$$

$$N=(\gamma\hat{})^{-1}X$$

REIM Formulation Cont.,

- Wage Equation

$$\log(Y_{i,t}/N_{i,t})=\alpha_0+\alpha_z(N_{i,t-1}/Y_{i,t-1})+\alpha_g G_{it}+\xi_{it}$$

- ✓ *Y is the income series and G includes compensation rate, hours worked, total production and unemployment rate.*

$$\mu_{it} = \exp(\alpha_0+\alpha_z(N_{i,t-1}/Y_{i,t-1})+\alpha_g G_{it})$$

$$Y=(\mu\hat{})N$$

Model Solution

- Inter-correlation of equation system
- Output Updating Equation

$$X = f(Y)$$

- Productivity Equation

$$N = f(X)$$

- Wage Equation

$$Y = f(N)$$

Commodity Flow Analysis

- Recovering I-O Coefficient (A) Matrix and Leontief Inverse (B) Matrix
- Estimating Inter-regional Commodity Flow Data

Recovering A and B Matrix

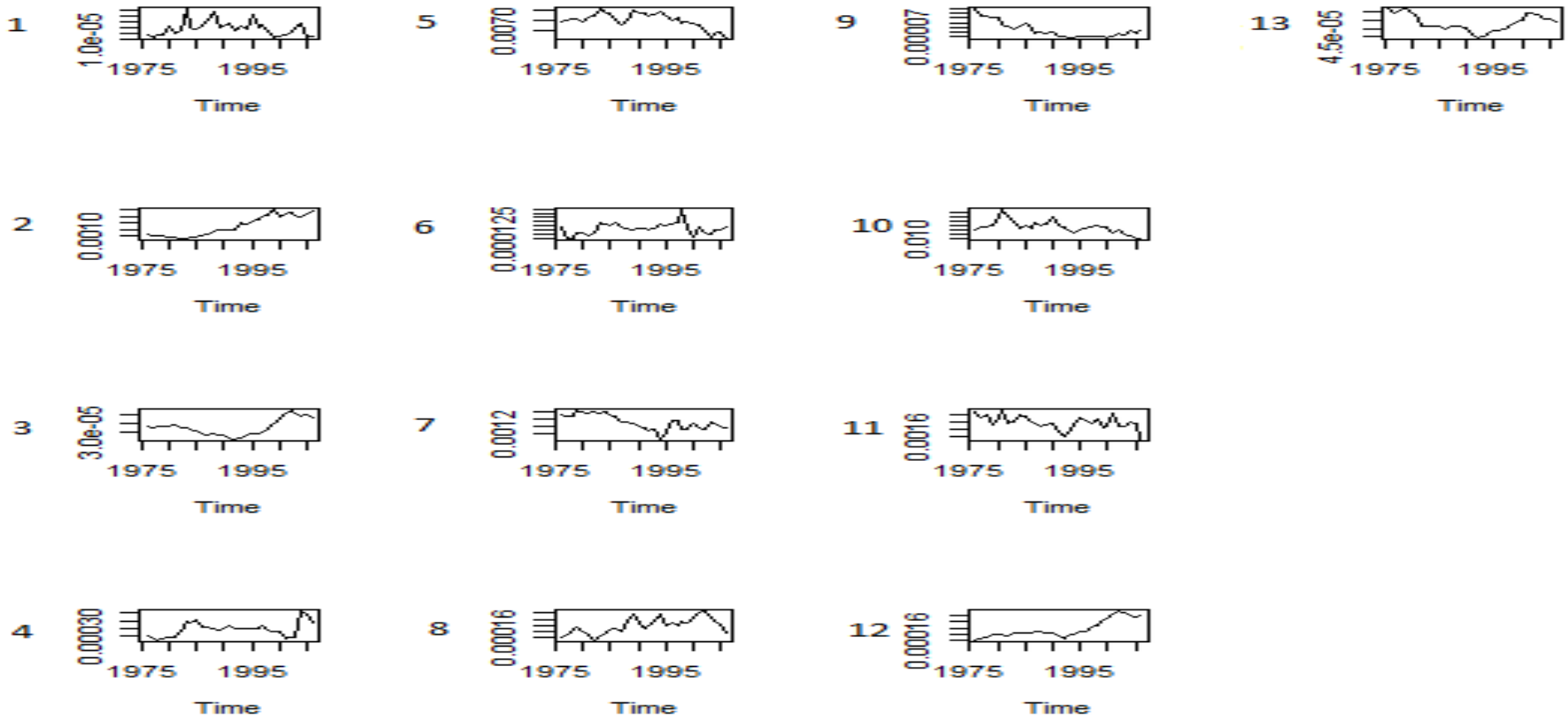
- General Procedure
- ✓ A similar practice will be adapted from Israilevich et al (1997) to recover the Leontief Inverse, with each element as output multiplier.
- ✓ An annual matrix could be recovered through the following relationship:

$$Bt=(I-At)^{-1}$$

Commodity Flow: At*X

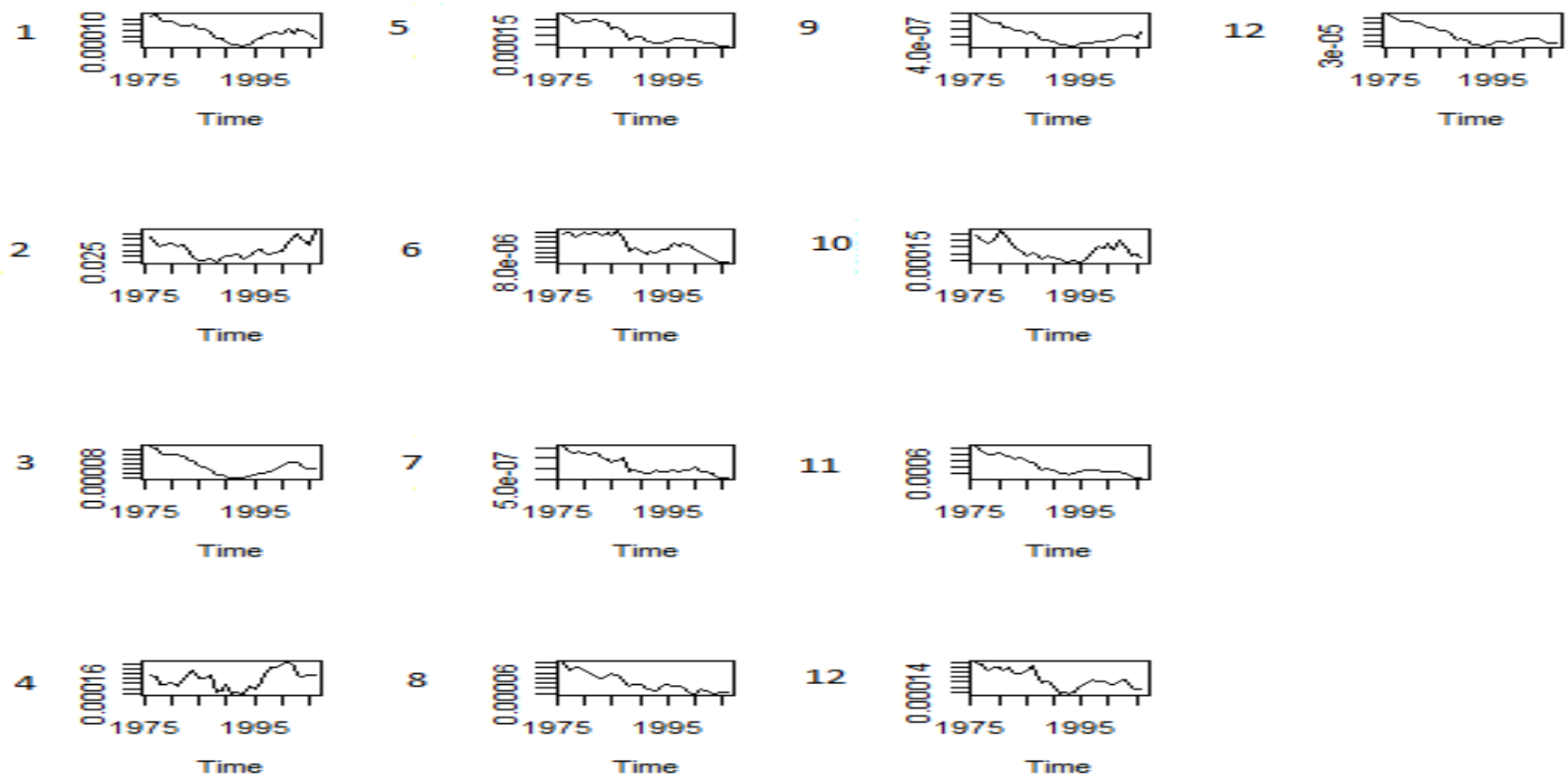
- ✓ The Value flows can be converted from dollars to tons using industry conversion factors.

Sample Results



Sales coefficients for sales from Chemical Manufacturing in Ohio to all industries in Illinois

Sample Results Cont.,



Sales coefficients for sales from Mining in Pennsylvania to all industries in New York

Sample Results, Cont

- No set scenarios discovered in the trade patterns in the industries considered.
- It remains to validate the data generated against published sources for specific years.
- After data have been validated, they will be made available to other scholars.
- We are now proceeding to estimation of the structural equation commodity-flow models.

Conclusion

Application of the data generated:

- ✓ Estimating dynamic structural-equation behavioral models;
- ✓ Impact analysis on infrastructure, air quality;
- ✓ Conducting thought experiments about appropriate policy responses to increasing trade.

Thank you for your attention☺

Any suggestions or comments will be greatly appreciated.