Estimating Metropolitan Exports

Nick Marchio

Background

- My project is an offshoot of an existing research project pertaining to metropolitan exports. The research aims to generate export estimates for over 130 goods and services industries over time for each of the 3,113 counties in the U.S. The resulting data is then provided to local economic development departments interested in devising plans that foster export-oriented growth.

- Due to limitations that exist in federal data there are no available sources of subnational exports that reflect actual sources of production outflows.

- Given this data deficit, traditional data science methods that rely on predictions based on samples or training data are not as helpful and we are forced to explore other techniques outside the traditional taxonomy of models.

	Continuous	Categorical
Supervised:	Regression	classification
Unsupervised:	dimension reduction	clustering

My Project

- One available solution is called allocation, a method that involves distributing a known control total available at higher levels of aggregation down to more detailed sub-levels using shares calculated from some other data source that strongly correlates with the data we're trying to estimate.

- Where this process gets complicated is if you introduce several different control totals at various levels of aggregation. This essentially creates a constraint satisfaction problem since all solutions must sum to the various control totals. Think of this as basically a scaled up version of Sudoku.

- Fortunately, there are algorithms that yields estimates that will converge on a solution that satisfies all constraints imposed by the control totals.

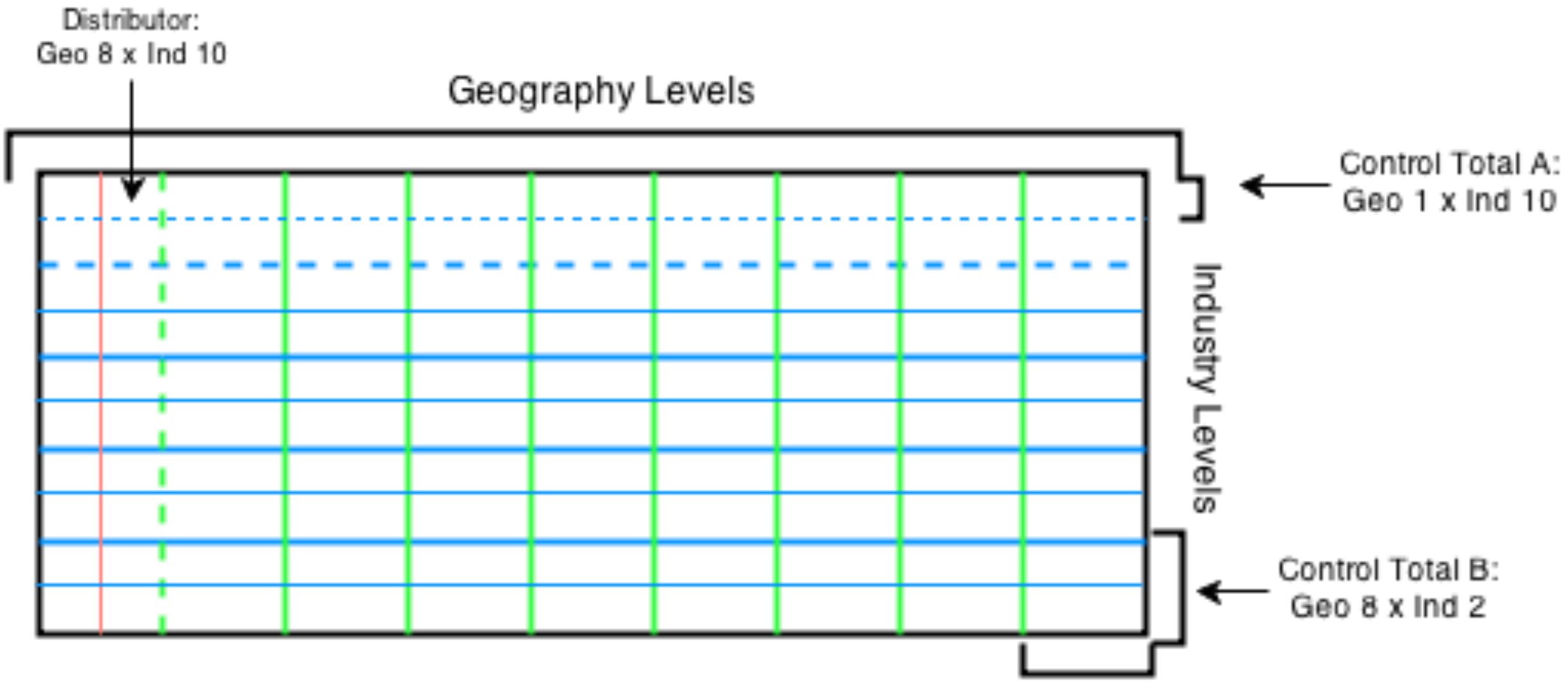
- Its key assumption is that all controls are mutually consistent and are structured in an orderly hierarchical manner so that it's possible to decompose totals into common subparts. In my data, I forced all the control totals to tie out with the national statistics and built non-overlapping schemas so that it would be possible to cleanly move between levels of aggregation.

- In the case of this project, there were a number of control totals including national detailed industry statistics and metropolitan statistics reported at a broader industry leve. The distributors were county GDP shares of each total. The control totals are then iteratively allocated, first to the shares of one total, and then to the shares of the other. This process is repeated, recalculating shares along the way, until the estimates converged on a set of mutually consistent solutions, thus satisfying the constraints.

Data

Function	Data Series	Year	Source
Goods Exports	NAICS 4 Domestic Goods Exports by U.S.	2003-2013	*Census USA Trade Online Goods Exports*
Service Exports	Services Exports by U.S.	2003-2013	*BEA Table 3.1 Service Exports*
	Detailed Services Exports by U.S.	2003-2012*	*BEA Table 1a Detailed Other Business Professional Technical*
	Detailed Travel & Tourism Spending by U.S.	2003-2013	*BEA Travel Tourism Satellite Industry Accounts*
	Detailed Royalties Industry Receipts by U.S.	2003-2011*	*IRS Royalties Table 7 Corporate Returns*
Inflation Adjustments	GDP Price Index by U.S.	2003-2013	*BEA Table 1.1.4 Price Indexes for Gross Domestic Product*
	Goods Export Price Index by U.S.	2003-2013	*BLS Producer Price Index*
	Agriculture & Services Export Price Index by U.S.	2003-2013	*BEA NIPA Table 4.2.4 Price Indexes for Exports and Imports of Goods and Services by Type of Product*
Spatial Allocation	NAICS 4 GDP by County	2003-2013	*Moody's Analytics*
	NAICS 3241 Exports by Petrol. Admin. Defense Districts (PADDs)	2003-2013	*Energy Information Administration's Total Crude Oil and Petroleum Products Exports*
	Intl. Origin, Destination, Passthrough Flights by Airport	2003, 2010*	*Sabre analysis of U.S. Department of Transportation's Bureau of Transportation Statistics International Flight Data*
	International Student Spending by Institution	2003-2013	*NAFSA analysis of IIE's Open Doors*
	SCTG 2 Exports by MSA and non-MSA	2010*	*EDR analysis of U.S. Department of Transportation's Freight Analysis Framwork, IMPLAN, WISER Trade, Oak Ridge Intercounty Impedances*
Jobs Estimation	NAICS 3-4 Direct & Total Multipliers by U.S.	2003-2012*	*BLS Employment Requirements Matrix*
	NAICS 3-4 Direct & Total Tourism-related Emp. by U.S.	2003-2013	*BEA Travel Tourism Satellite Industry Accounts*

Diagram


Distributor:
Geo 8 x Ind 10
Geography Levels
Control Total A:
Geo 1 x Ind 10
Industry Levels
Control Total B:
Geo 8 x Ind 2

Algebra

$$\hat{x}_i = X_{control} * \left({z_i}/{\sum_{i=1}^{n} z_j} \right)$$

$\hat{x}_i$ = estimate for the component part of $X_{control}$ by county & detailed industry

$X_{control}$ = given control total that can be subdivided by county & industry

z_i = the county & detailed industry value of Z data, being used to allocate the value of $X_{control}$

$\sum_{i=1}^{n} z_j = z_i$ values aggregated according to the groups available in the Z data

n = number of groups (counties * industries)

Pseudocode

In terms of pseudocode, the algorithm may be implemented as follows:

For $\boldsymbol{i}$ in X-Data iterate until $\boldsymbol{a_i}$ ***and*** $\boldsymbol{b_i}$ converge:

$Sum\text{-}A = \sum_{i=1}^{nA} \boldsymbol{a_j}$ = Aggregate $\boldsymbol{a_i}$ by j groupings in $\boldsymbol{Xa_{control}}$

$Share\text{-}A = {}^{\boldsymbol{a_i}}/_{\sum_{i=1}^{n} \boldsymbol{a_j}}$ = Calculate $\boldsymbol{a_i}$ share of $\sum_{i=1}^{n} \boldsymbol{a_j}$

$\boldsymbol{\hat{x}_i}$ = Allocate $\boldsymbol{Xa_{control}}$ to *Share-A*

$\boldsymbol{b_i} = \boldsymbol{\hat{x}_i}$

$Sum\text{-}B = \sum_{i=1}^{nA} \boldsymbol{b_j}$ = Aggregate $\boldsymbol{b_i}$ by j groupings in $\boldsymbol{Xb_{control}}$

$Share\text{-}B = {}^{\boldsymbol{b_i}}\Big/_{\sum_{i=1}^{n} \boldsymbol{b_j}}$ = Calculate $\boldsymbol{b_i}$ share of $\sum_{i=1}^{n} \boldsymbol{b_j}$

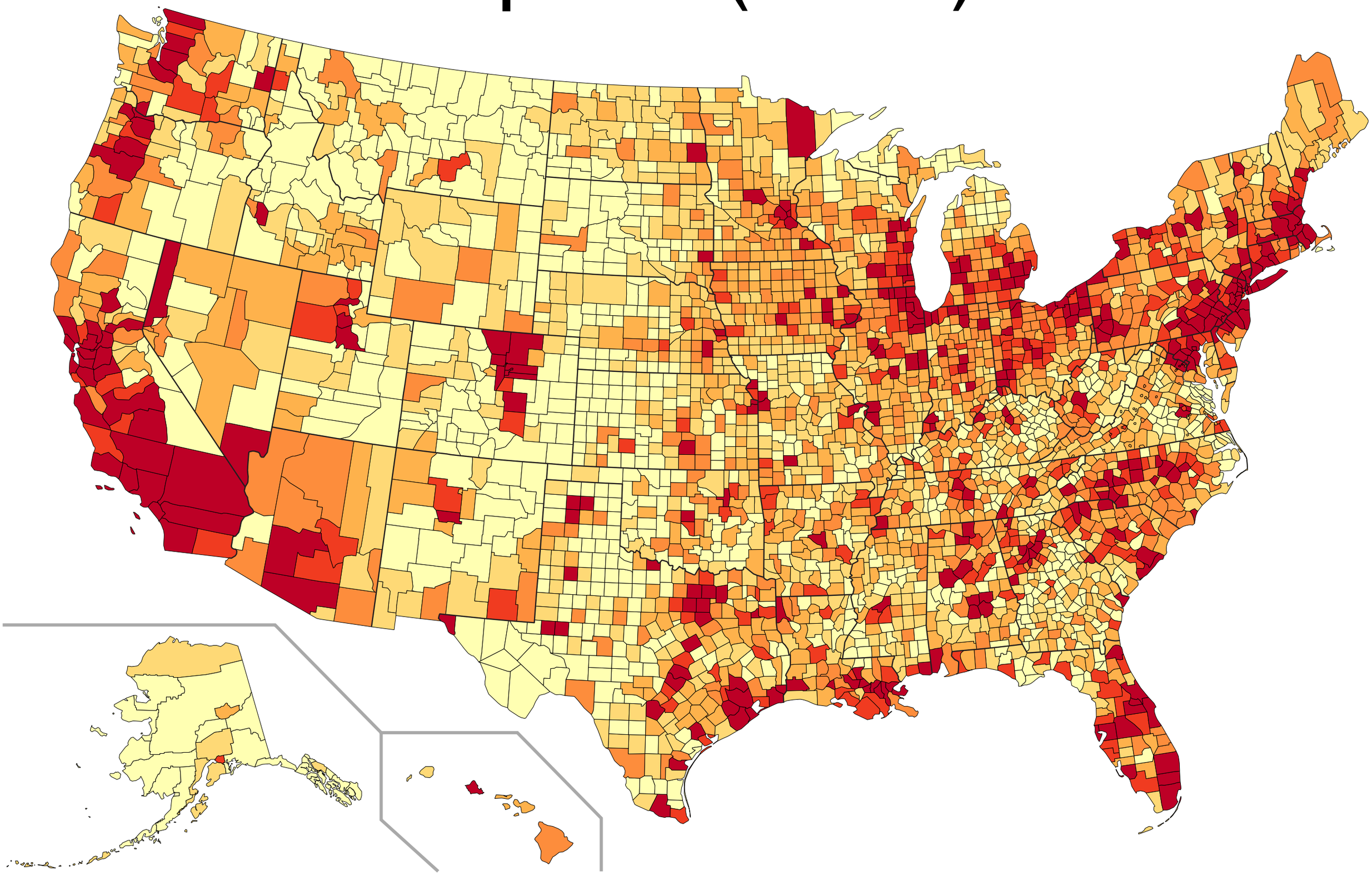
$\boldsymbol{\hat{x}_i}$ = Allocate $\boldsymbol{Xb_{control}}$ to *Share-B*

$\boldsymbol{a_i} = \boldsymbol{\hat{x}_i}$

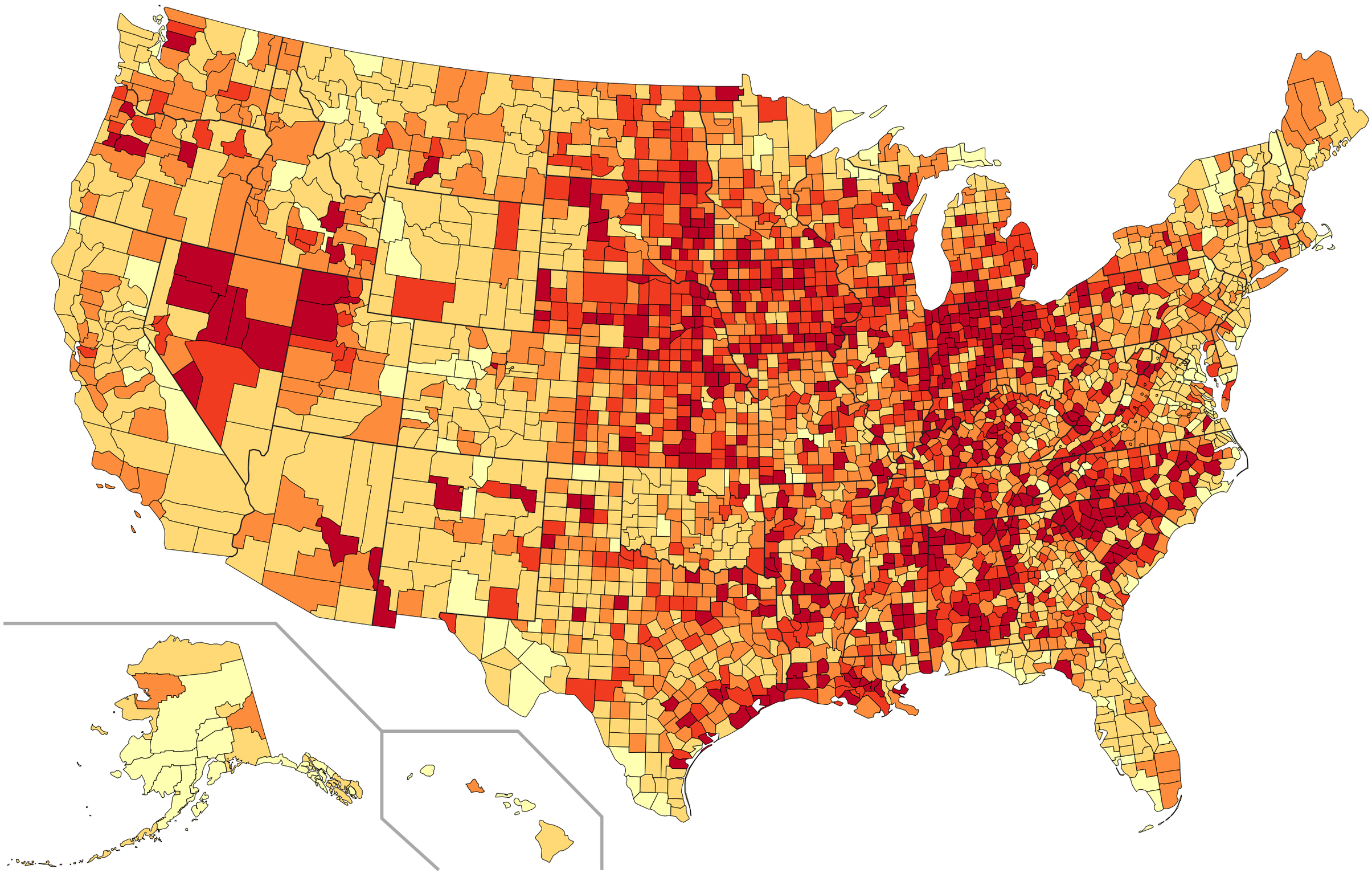
Applications

- The main takeaway is that this technique is a handy way to derive more detailed estimates from aggregate data as long as you have other measures that you think correlate with the data you're trying to estimate.
- It is attractive because it easily allows you to fit your data within the framework of other more reliable control data sets.
- Believe it or not, this algorithm is used by a wide array of data organizations.
- BEA uses it to estimate state level estimates of GDP using compensation data obtained from the IRS.
- Moody's Analytics also uses the technique to allocate state level GDP to the county level using the QCEW compensation data from the BLS.

Exports (2013)

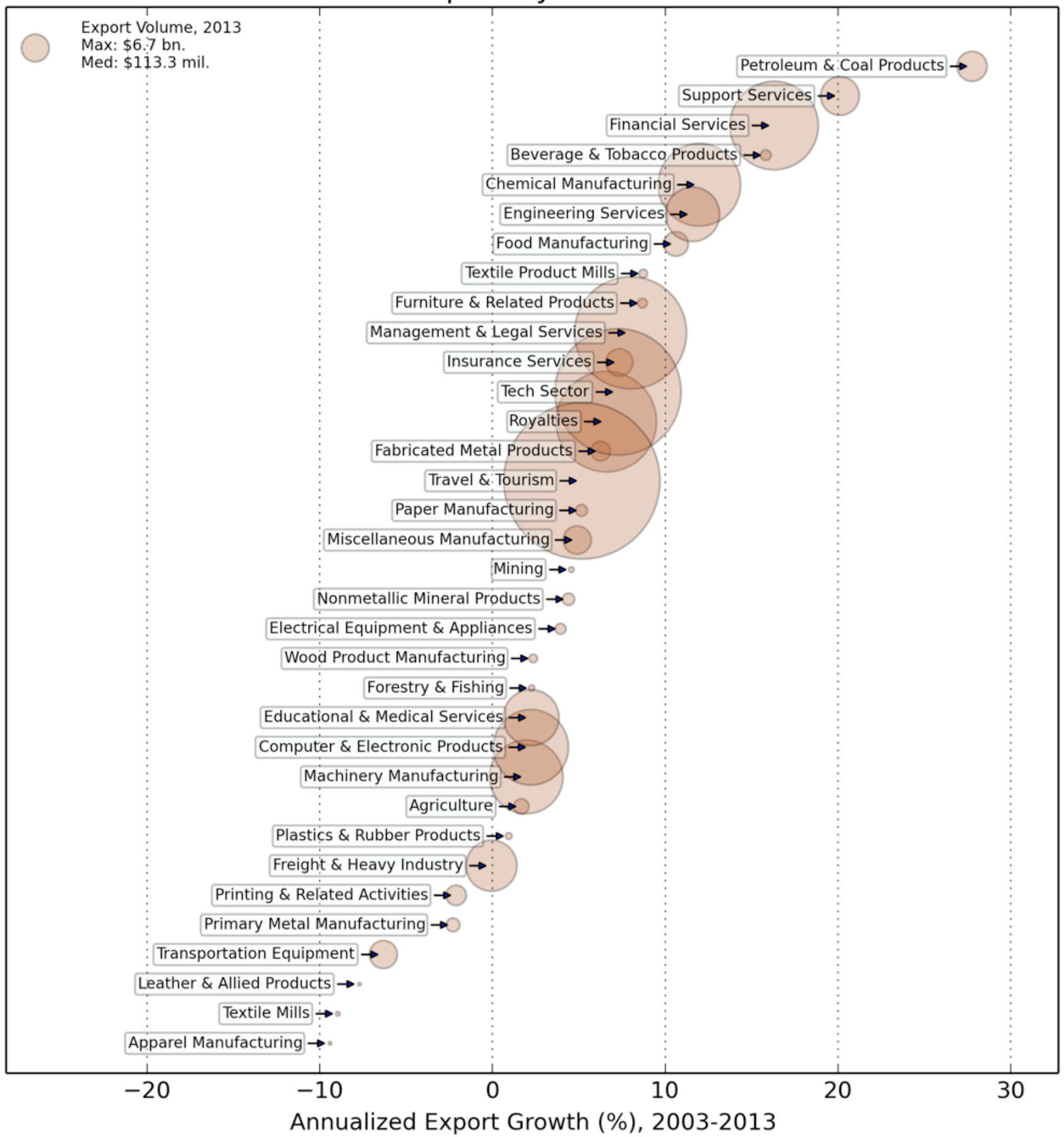


Export Share of GDP (2013)

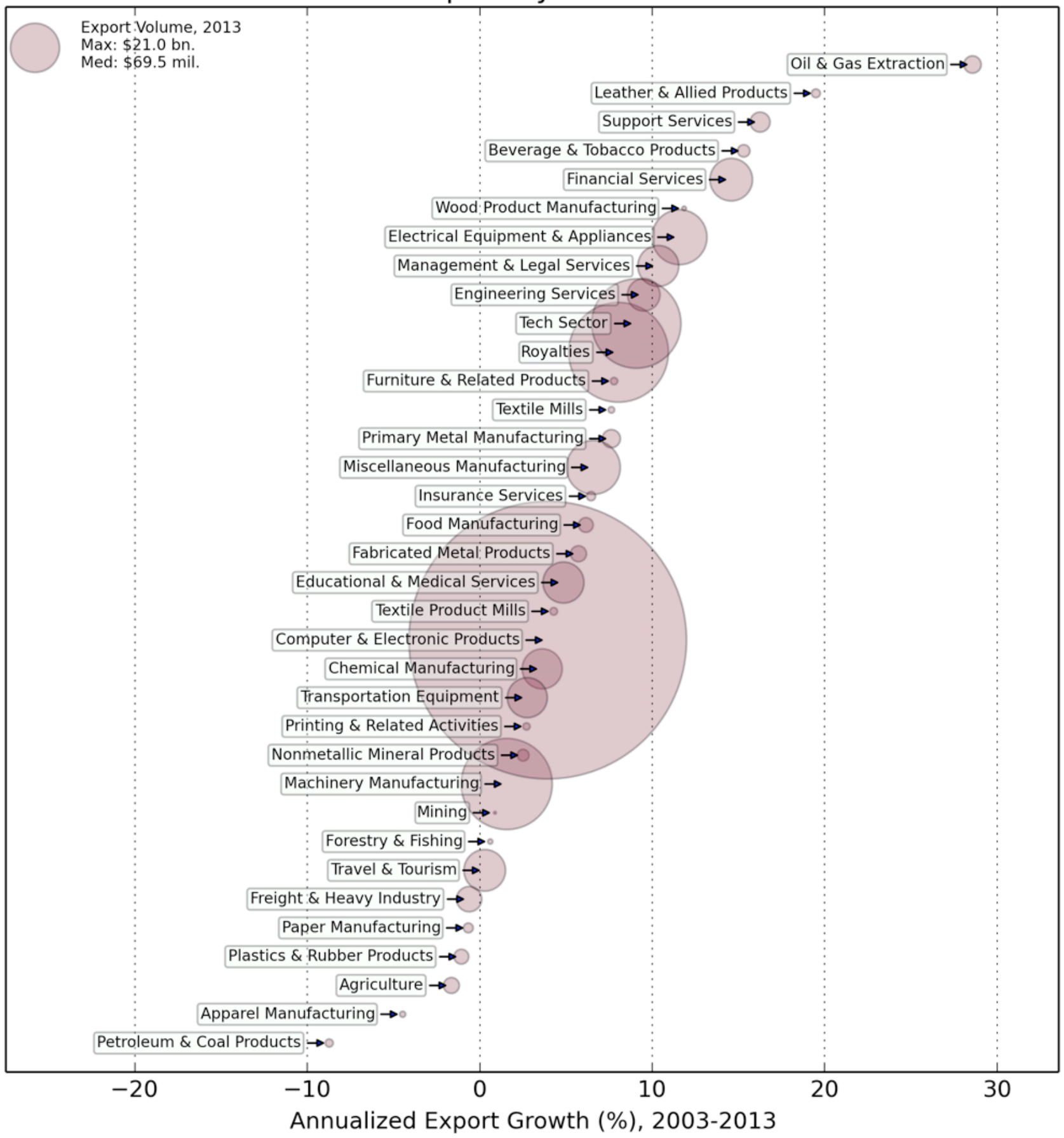


Tuesday, September 30, 14

Washington, DC-VA-MD:
Export Dynamics



San Jose, CA:
Export Dynamics



New York, NY-NJ-PA: Export Dynamics