RELAX, TENSORS ARE HERE...WITH EXOGENOUS COVARIATES

Peter D. Hoff, Shahryar Minhas, & Michael D. Ward

June 29, 2015

model specification

Dependent variables: Log(Exports) and Stdzed(Material Conflict).

Direct (*i*), reciprocal (*ji*) and transitive (*ijk*) 1 month lags of these included as IVs.

Exogenous Covariates:

- Number of Preferential Trade Agreements (PTA) between *i* and *j* (this is an undirected, yearly level variable). Direct and transitive version of this variable included as covariates.
- Presence of a defensive alliance relationship between *i* and *j* (undirected, yearly level). Direct and transitive versions.
- Centroid distance between *i* and *j* (directed). Direct version.
- Polity, monthly level variable. Polity of sender included.
- Log(GDP), yearly level variable but imputed at the monthly level. GDP of sender.
- Log(Population), yearly level variable but imputed at the monthly level. Population of sender.
- Log(Total Exports to any country), monthly level variable. Exports of sender.

sample & data

Our sample is comprised of 161 countries over the period of March 2001 to December 2014

Data sources:

- Exports: IMF Direction of Trade Statistics
- Material Conflict: ICEWS
- PTA: Design of Trade Agreements Database
- Alliance: Correlates of War
- Distance: cshapes
- Polity: Polity IV Project
- GDP, Population: IMF World Economic Outlook Database

modeling approach

Multilinear tensor regression framework

MCMC run for 1300 iterations with first 600 used as burn-in[1]

The model has the following form:

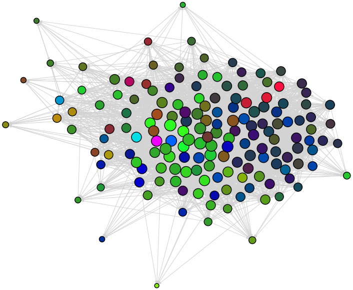
$$\mathbf{Y} = \mathbf{X} \times \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3\} + \mathbf{E}$$

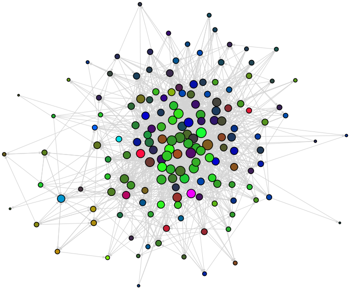
Y is a $161 \times 161 \times 2 \times 165$ array

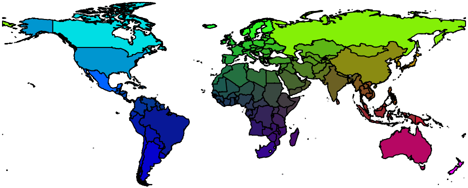
X is a $161 \times 161 \times 13 \times 165$ array, where each of the 13 variables is lagged by one month

[1] Using this many datapoints takes time the MCMC will keep running for another 3700 iterations so these results are preliminary, but trace plots at the end of this pdf look stable after 600 iterations

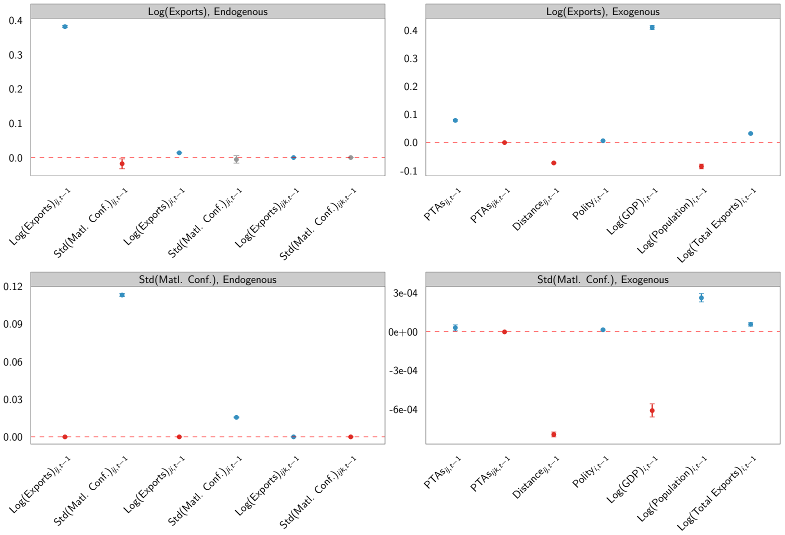
β_1 & β_2, sig. + shown, $\alpha = 0.01$







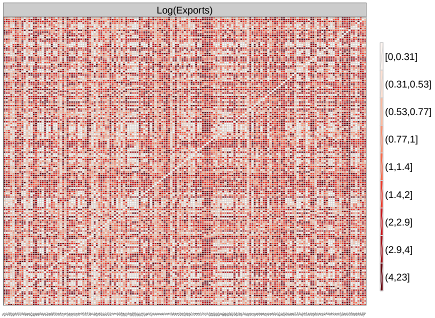
β_3

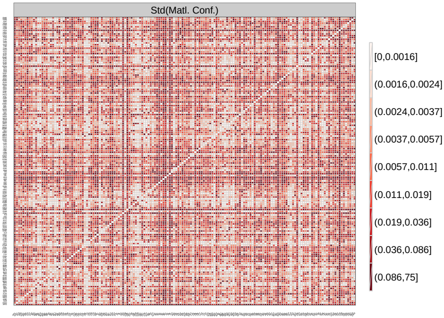


Log(Exports), Endogenous
Log(Exports), Exogenous
Std(Matl. Conf.), Endogenous
Std(Matl. Conf.), Exogenous
0.4
0.3
0.2
0.1
0.0
-0.1
0.12
0.09
0.06
0.03
0.00
3e-04
0e+00
-3e-04
-6e-04
$\mathrm{Log(Exports)}_{ij,t-1}$
$\mathrm{Std(Matl.\ Conf.)}_{ij,t-1}$
$\mathrm{Log(Exports)}_{ji,t-1}$
$\mathrm{Std(Matl.\ Conf.)}_{ji,t-1}$
$\mathrm{Log(Exports)}_{ijk,t-1}$
$\mathrm{Std(Matl.\ Conf.)}_{ijk,t-1}$
$\mathrm{PTAs}_{ij,t-1}$
$\mathrm{PTAs}_{ijk,t-1}$
$\mathrm{Distance}_{ij,t-1}$
$\mathrm{Polity}_{i,t-1}$
$\mathrm{Log(GDP)}_{i,t-1}$
$\mathrm{Log(Population)}_{i,t-1}$
$\mathrm{Log(Total\ Exports)}_{i,t-1}$

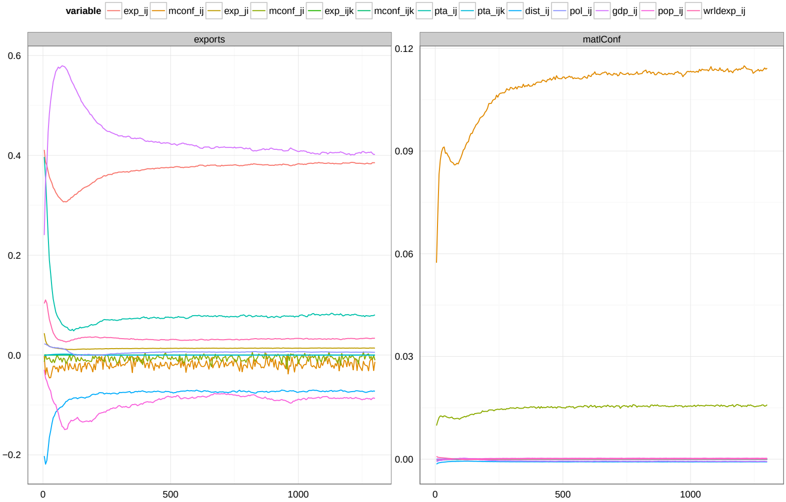
aggregate performance & rmse by i-j

	RMSE	R^2
Log(Exports)	2.32	0.95
Std(Matl. Conf.)	0.85	0.28





trace plots for β_3



comparison with directed dyadic model

Here I run a similar analysis using the standard directed dyadic (dd) framework

The covariates for both models are the same

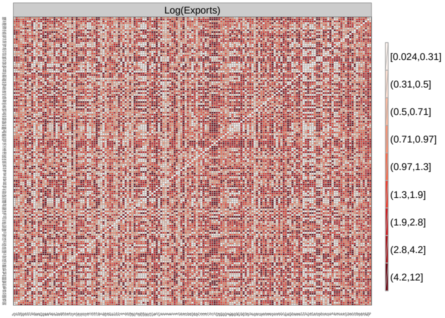
Instead of taking a vector autoregression approach, I just run two separate directed dyadic linear regressions

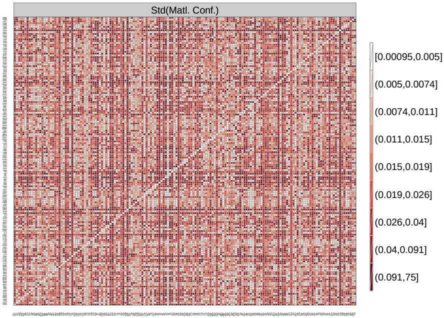
dd coefficient results, std. errors in (), * sig. at $p < 0.05$

	Log(Exports)	Std(Matl. Conf.)
(Intercept)	0.20^*	0.22^*
	(0.02)	(0.01)
exp_ij	0.82^*	-0.00^*
	(0.00)	(0.00)
mconf_ij	-0.00	0.49^*
	(0.00)	(0.00)
exp_ji	0.06^*	-0.00
	(0.00)	(0.00)
mconf_ji	0.00	0.03^*
	(0.00)	(0.00)
exp_ijk	0.00^*	0.00^*
	(0.00)	(0.00)
mconf_ijk	-0.00^*	0.00^*
	(0.00)	(0.00)
pta_ij	0.10^*	-0.00
	(0.00)	(0.00)
pta_ijk	-0.00^*	-0.00^*
	(0.00)	(0.00)
dist_ij	-0.09^*	-0.01^*
	(0.00)	(0.00)
pol_ij	0.01^*	-0.00^*
	(0.00)	(0.00)
gdp_ij	0.13^*	0.01^*
	(0.00)	(0.00)
pop_ij	-0.02^*	0.00^*
	(0.00)	(0.00)
wrldexp_ij	0.01^*	-0.01^*
	(0.00)	(0.00)
N	4250400	4250400

dd aggregate performance & rmse by i-j

	RMSE	R^2
Log(Exports)	2.37	0.89
Std(Matl. Conf.)	0.86	0.26





Next Steps?