‘MATH+ECON+CODE’ MASTERCLASS ON MATCHING MODELS, OPTIMAL TRANSPORT AND APPLICATIONS

Alfred Galichon (New York University)

Spring 2018
Day 5, January 19 2018: “Empirical matching models”
Block 14. The gravity equation

Learning objectives: block 14

- regularized optimal transport
- the gravity equation
- generalized linear models
- pseudo-Poisson maximum likelihood estimation

- Anderson and van Wincoop (2003). "Gravity with Gravitas: A Solution to the Border Puzzle". *AER*.
- Head and Mayer (2014). "Gravity Equations: Workhorse, Toolkit and Cookbook". *Handbook of international economics.*
- Gourieroux, Trognon, Monfort (1984). "Pseudo Maximum Likelihood Methods: theory" *Econometrica*.
- McCullagh and Nelder (1989). *Generalized Linear Models*. Chapman and Hall/CRC.
- Santos Silva and Tenreyro (2006). "The Log of Gravity". *REStats*.
- Yotov et al. (2011). *An advanced guide to trade policy analysis*. WTO.
- Guimares and Portugal (2012). "Real Wages and the Business Cycle: Accounting for Worker, Firm, and Job Title Heterogeneity". *AEJ: Macro*.
- Dupuy and G (2014), "Personality traits and the marriage market". *JPE*.
- Dupuy, G and Sun (2016), "Estimating matching affinity matrix under low-rank constraints." arxiv 1612.09585.

Section 1

THEORY

- The gravity equation is a very useful tool for explaining trade flows by various measures of proximity between countries.
- A number of regressors have been proposed. They include: geographic distance, common official languague, common colonial past, share of common religions, etc.
- The dependent variable is the volume of exports from country i to country n, for each pair of country (i, n).
- Today, we shall see a close connection between gravity models of international trade and separable matching models.

- Consider the optimal transport duality

$$\max_{\pi \in \mathcal{M}(P,Q)} \sum_{xy} \pi_{xy} \Phi_{xy} = \min_{u_x + v_y \geq \Phi_{xy}} \sum_{x \in \mathcal{X}} p_x u_x + \sum_{y \in \mathcal{Y}} q_y v_y$$

- Now let's assume that we are adding an entropy to the primal objective function. For any $\sigma > 0$, we get

$$\begin{aligned} &\max_{\pi \in \mathcal{M}(P,Q)} \sum_{xy} \pi_{xy} \Phi_{xy} - \sigma \sum_{xy} \pi_{xy} \ln \pi_{xy} \\ &= \min_{u,v} \sum_{x \in \mathcal{X}} p_x u_x + \sum_{y \in \mathcal{Y}} q_y v_y + \sigma \sum_{xy} \exp\left(\frac{\Phi_{xy} - u_x - v_y - \sigma}{\sigma} \right) \end{aligned}$$

- The latter problem is an unconstrained convex optimization problem. But the most efficient numerical computation technique is often coordinate descent, i.e. alternate between minimization in u and minimization in v.

- Maximize wrt to u yields

$$e^{-u_x/\sigma} = \frac{p_x}{\sum_y \exp\left(\frac{\Phi_{xy}-v_y-\sigma}{\sigma}\right)}$$

and wrt v yields

$$e^{-v_y/\sigma} = \frac{q_y}{\sum_x \exp\left(\frac{\Phi_{xy}-v_y-\sigma}{\sigma}\right)}$$

- It is called the "iterated projection fitting procedure" (ipfp), aka "matrix scaling", "RAS algorithm", "Sinkhorn-Knopp algorithm", "Kruithof's method", "Furness procedure", "biproportional fitting procedure", "Bregman's procedure". See survey in Idel (2016).
- Maybe the most often reinvented algorithm in applied mathematics. Recently rediscovered in a machine learning context.

- The goal is to estimate the matching surplus Φ_{xy}. For this, take a linear parameterization

$$\Phi_{xy}^{\beta} = \sum_{k=1}^{K} \beta_k \phi_{xy}^{k}.$$

- Following Choo and Siow (2006), G and Salanié (2017) introduce logit heterogeneity in individual preferences and show that the equilibrium now maximizes the *regularized Monge-Kantorovich problem*

$$W(\beta) = \max_{\pi \in \mathcal{M}(P,Q)} \sum_{xy} \pi_{xy} \Phi_{xy}^{\beta} - \sigma \sum_{xy} \pi_{xy} \ln \pi_{xy}$$

- By duality, $W(\beta)$ can be expressed

$$W(\beta) = \min_{u,v} \sum_{x} p_x u_x + \sum_{y} q_y v_y + \sigma \sum_{xy} \exp\left(\frac{\Phi_{xy}^{\beta} - u_x - v_y - \sigma}{\sigma} \right)$$

and w.l.o.g. can set $\sigma = 1$ and drop the additive constant $-\sigma$ in the exp.

- We observe the actual matching $\hat{\pi}_{xy}$. Note that $\partial W / \partial \beta^k = \sum_{xy} \pi_{xy} \phi^k_{xy}$, hence β is estimated by running

$$\min_{u,v,\beta} \sum_x p_x u_x + \sum_y q_y v_y + \sum_{xy} \exp\left(\Phi^\beta_{xy} - u_x - v_y \right) - \sum_{xy,k} \hat{\pi}_{xy} \beta_k \phi^k_{xy} \tag{1}$$

  which is still a convex optimization problem.
- This is actually the objective function of the log-likelihood in a Poisson regression with x and y fixed effects, where we assume

$$\pi_{xy} | xy \sim Poisson\left(\exp\left(\sum_{k=1}^K \beta_k \phi^k_{xy} - u_x - v_y \right) \right).$$

- Let $\theta = (\beta, u, v)$ and $Z = (\phi, D^x, D^y)$ where $D^x_{x'y'} = 1\{x = x'\}$ and $D^y_{x'y'} = 1\{y = y'\}$ are x-and y-dummies. Let $m_{xy}(Z;\theta) = \exp(\theta^\intercal Z_{xy})$ be the parameter of the Poisson distribution.
- The conditional likelihood of $\hat{\pi}_{xy}$ given Z_{xy} is

$$\begin{aligned} l_{xy}(\hat{\pi}_{xy};\theta) &= \hat{\pi}_{xy} \log m_{xy}(Z;\theta) - m_{xy}(Z;\theta) \\ &= \hat{\pi}_{xy}(\theta^\intercal Z_{xy}) - \exp(\theta^\intercal Z_{xy}) \\ &= \hat{\pi}_{xy}\left(\sum_{k=1}^K \beta_k \phi^k_{xy} - u_x - v_y\right) - \exp\left(\sum_{k=1}^K \beta_k \phi^k_{xy} - u_x - v_y\right) \end{aligned}$$

- Summing over x and y, the sample log-likelihood is

$$\sum_{xy} \hat{\pi}_{xy} \sum_{k=1}^K \beta_k \phi^k_{xy} - \sum_x p_x u_x - \sum_y q_y v_y - \sum_{xy} \exp\left(\sum_{k=1}^K \beta_k \phi^k_{xy} - u_x - v_y\right)$$

hence we recover objective function (1).

- If $\pi_{xy}|xy$ is Poisson, then $\mathbb{E}\left[\pi_{xy}\right] = m_{xy}\left(Z_{xy};\theta\right) = \mathbb{V}ar\left(\pi_{xy}\right)$. While it makes sense to assume the former equality, the latter is a rather strong assumption.
- For estimation purposes, $\hat{\theta}$ is obtained by

$$\max_{\theta} \sum_{xy} l\left(\hat{\pi}_{xy};\theta\right) = \sum_{xy}\left(\hat{\pi}_{xy}\left(\theta^{\intercal} Z_{xy}\right) - \exp\left(\theta^{\intercal} Z_{xy}\right)\right)$$

however, for inference purposes, one shall not assume the Poisson distribution. Instead

$$\sqrt{N}\left(\hat{\theta}-\theta\right) \Longrightarrow \left(A_0\right)^{-1} B_0 \left(A_0\right)^{-1}$$

where $N = |\mathcal{X}| \times |\mathcal{Y}|$ and A_0 and B_0 are estimated by

$$\hat{A}_0 = N^{-1}\sum_{xy} D^2_{\theta\theta} l\left(\hat{\pi}_{xy};\hat{\theta}\right) = N^{-1}\sum_{xy}\exp\left(\hat{\theta}^{\intercal} Z_{xy}\right) Z_{xy} Z_{xy}^{\intercal}$$

$$\hat{B}_0 = N^{-1}\sum_{xy}\left(\hat{\pi}_{xy} - \exp\left(\hat{\theta}^{\intercal} Z_{xy}\right)\right)^2 Z_{xy} Z_{xy}^{\intercal}.$$

- Dupuy and G (2014) focus on cross-dimensional interactions

$$\phi_{xy}^{A} = \sum_{p,q} A_{pq} \xi_x^p \xi_y^q$$

  and estimate "affinity matrix" A on a dataset of married individuals where the "big 5" personality traits are measured.
- A is estimated by

$$\min_{s_i, m_n} \min_{A} \left\{ \begin{array}{c} \sum_x p_x u_x + \sum_y q_y v_y \\ + \sum_{xy} \exp\left(\sum_{p,q} A_{pq} \xi_x^p \xi_y^q - u_x - v_y \right) \\ - \sum_{x,y,p,q} \hat{\pi}_{xy} A_{pq} \xi_x^p \xi_y^q \end{array} \right\}.$$

- Dupuy, G and Sun (2016) consider the case when the space of characteristics is high-dimensional. More on this this afternoon.

TABLE: Affinity matrix. Source: Dupuy and G (2014).

Wives / Husbands	Education	Height.	BMI	Health	Consc.	Extra.	Agree.	Emotio.	Auto.	Risk
Education	**0.46**	0.00	**-0.06**	0.01	-0.02	0.03	-0.01	-0.03	0.04	0.01
Height	0.04	**0.21**	0.04	0.03	**-0.06**	0.03	0.02	0.00	-0.01	0.02
BMI	-0.03	0.03	**0.21**	0.01	0.03	0.00	**-0.05**	0.02	0.01	-0.02
Health	-0.02	0.02	- 0.04	**0.17**	- 0.04	0.02	- 0.01	0.01	-0.00	0.03
Conscienciousness	**-0.07**	-0.01	**0.07**	-0.00	**0.16**	0.05	0.04	0.06	0.01	0.01
Extraversion	0.00	-0.01	0.00	0.01	**-0.06**	**0.08**	-0.04	-0.01	0.02	**-0.06**
Agreeableness	0.01	0.01	-0.06	0.02	**0.10**	**-0.11**	0.00	0.07	-0.07	-0.05
Emotional	0.03	-0.01	0.04	**0.06**	**0.19**	0.04	0.01	-0.04	**0.08**	**0.05**
Autonomy	0.03	0.02	0.01	0.02	**-0.09**	**0.09**	-0.04	0.02	**-0.10**	0.03
Risk	0.03	-0.01	-0.03	-0.01	0.00	-0.02	-0.03	-0.03	**0.08**	**0.14**

Note: Bold coefficients are significant at the 5 percent level.

- "Structural gravity equation" (Anderson and van Wincoop, 2003) as reviewed in Head and Mayer (2014) handbook chapter:

$$X_{ni} = \underbrace{\frac{Y_i}{\Omega_i}}_{S_i} \underbrace{\frac{X_n}{\Psi_n}}_{M_n} \Phi_{ni}$$

  where n=importer, i=exporter, X_{ni}=trade flow from i to n, $Y_i = \sum_n X_{ni}$ is value of production, $X_n = \sum_i X_{ni}$ is importers' expenditures, and ϕ_{ni}=bilateral accessibility of n to i.

- Ω_i and Ψ_n are "multilateral resistances", satisfying the set of implicit equations

$$\Psi_n = \sum_i \frac{\Phi_{ni} Y_i}{\Omega_i} \text{ and } \Omega_i = \sum_n \frac{\Phi_{ni} X_n}{\Psi_n}$$

- These are exactly the same equations as those of the regularized OT.

- Parameterize $\Phi_{ni} = \exp\left(\sum_{k=1}^{K} \beta_k D_{ni}^k\right)$, where the D_{ni}^k are K pairwise measures of distance between n and i. We have

$$X_{ni} = \exp\left(\sum_{k=1}^{K} \beta_k D_{ni}^k - s_i - m_n\right)$$

where fixed effects $s_i = -\ln S_i$ and $m_n = -\ln M_n$ are adjusted by

$$\sum_i X_{ni} = Y_i \text{ and } \sum_n X_{ni} = X_n.$$

- Standard choices of D_{ni}^k's:
  - logarithm of bilateral distance between n and i
  - indicator of contiguous borders; of common official language; of colonial ties
  - trade policy variables: presence of a regional trade agreement; tariffs
  - could include many other measures of proximity, e.g. measure of genetic/cultural distance, intensity of communications, etc.

Section 2

CODING

- We replicate table 1 p. 42 of Yotov et al. available fromhttps://www.wto.org/english/res_e/booksp_e/advancedwtounctad2016_e.p