# Network Econometrics: Unbiased Estimation of Extensive and Intensive Margins

Marzio Di Vece, Tiziano Squartini & Diego Garlaschelli

## Motivation

Can Maximum Entropy Network-based models improve Econometric Gravity Models ?
In the 2000s PPML [1] and Zero-Inflated models [2] have been proposed in Econometric literature to tackle respectively heteroscedasticty and the excess zeroes problems. Those econometric models do not provide a correct estimation of certain topological properties [3].
Our goal is to use a Maximum Entropy framework Network-based model that can take as input a general gravity specification and is more reliable in providing topological statistics. We improve on the Enhanced Gravity Model (EGM) [4], generating a set of innovative maximum entropy models that can serve as econometric models. We systematically study the performance of such models in terms of model selection measures and accuracy in reproducing higher-order topological and weighted quantities in an undirected framework.

## Model Selection

| Models | $\Delta L$ | $\Delta W$ |
|---|---|---|
| EGM-1 | $0^*$ | $0^*$ |
| EGM-2 | $0^*$ | $0^*$ |
| EGM-TS | $0^*$ | $\simeq 3.5 \cdot 10^{-3}$ |
| EGM-TSF | $0^*$ | $\simeq 1.3 \cdot 10^{-3}$ |

**Table 1:** Macro-Error: Network-based models. $0^*$ stands for zero if not for numeric error.

| Models | $\Delta L$ | $\Delta W$ |
|---|---|---|
| POIS | $\simeq 0.07$ | $0^*$ |
| ZIP | $\simeq 4.5 \cdot 10^{-3}$ | $\simeq 3.3 \cdot 10^{-3}$ |
| NB2 | $\simeq 0.06$ | $\simeq 0.59$ |
| ZINB | $\simeq 0.06$ | $\simeq 0.95$ |

**Table 2:** Macro-Error: Econometric-based models. ZIP is good in reproducing $L$ and $W$ but fails in predicting higher-order network statistics data variation.

| Models | AIC | BIC |
|---|---|---|
| EGM-1 | $\simeq 182909.5$ | $\simeq 182951.2$ |
| EGM-2 | $\simeq 174050.0$ | $\simeq 175550.4$ |
| EGM-TS | $\simeq 182516.7$ | $\simeq 184017.0$ |
| EGM-TSF | $\simeq 192327.7$ | $\simeq 192369.4$ |

**Table 3:** MS: Network-based models. EGM-2 performs better than purely econometric models according to AIC/BIC.

| Models | AIC | BIC |
|---|---|---|
| POIS | $\simeq 5088080.1$ | $\simeq 5088105.1$ |
| ZIP | $\simeq 5052767.3$ | $\simeq 5052800.6$ |
| NB2 | $\simeq 175693.2$ | $\simeq 175726.6$ |
| ZINB | $\simeq 176071.4$ | $\simeq 176113.1$ |

**Table 4:** MS: Econometric-based models. EGM-2 performs better than purely econometric models according to AIC/BIC.

## Network Statistics

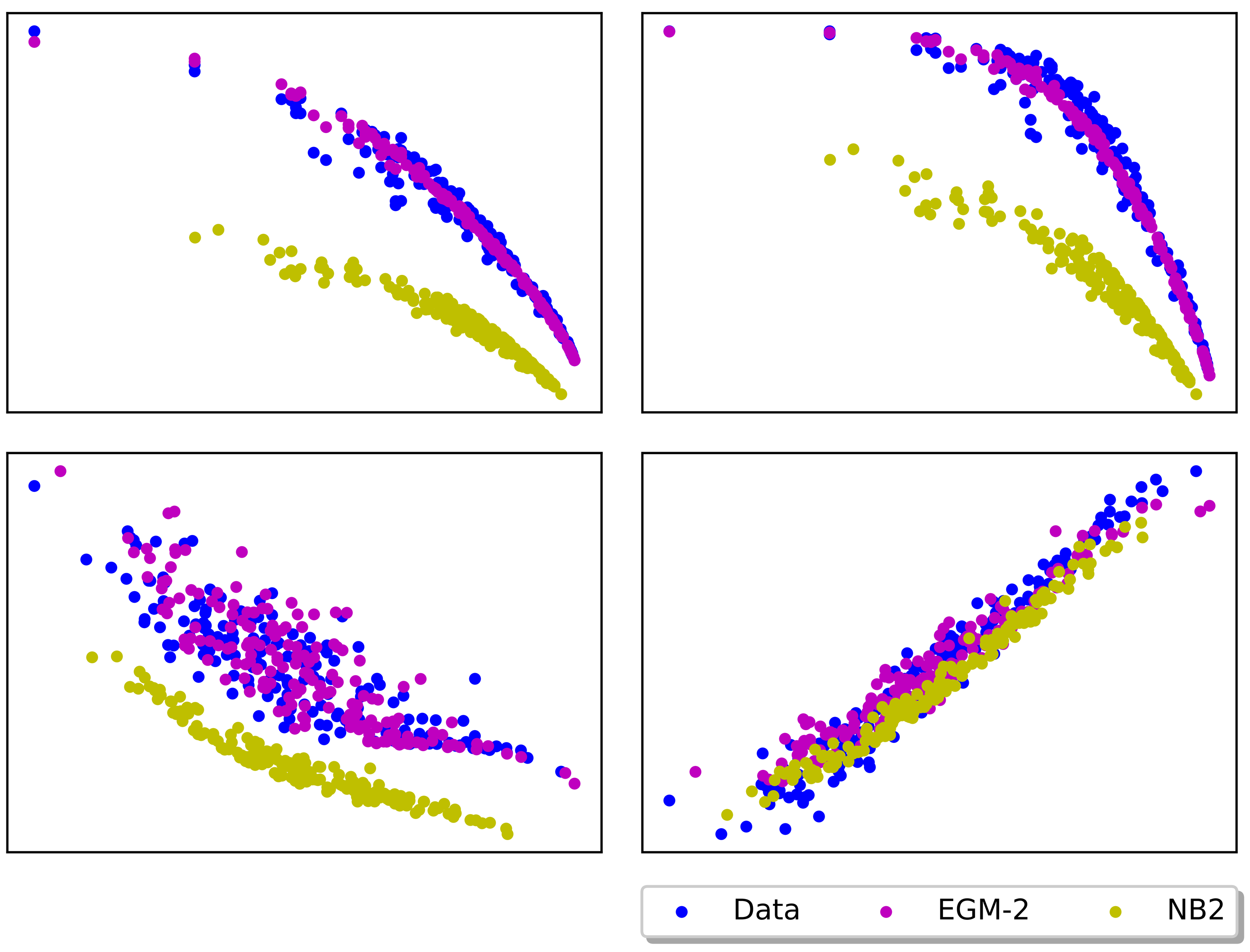


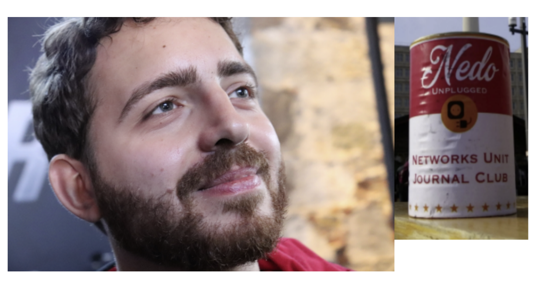

**Figure 1:** Comparison between EGM-2 and NB2 models: (top-left) average neighbors degree and (top-right) clustering coefficient against average degree, (bottom-left) average neighbor strength and (bottom-right) linear weighted clustering against average strength. Real data is depicted in blue, EGM-2 in magenta and NB2 in yellow. NB2, the best performing purely Econometric model according to AIC/BIC systematically underestimates average neighbor degree, clustering and average neighbor strength statistics. EGM-2, instead, performs very well reproducing data variation and trends in all of the key statistics.

## Main References

[1] J Santos Silva et al. *Rev. Econ. Stat. E* *88(4)*, 2006.
[2] M Burger et al. *Spat. Econ. Anal.* *4(2)*, 2009.
[3] M Duenas et al. *J. Econ. Interact. Coord.* **8**, 2013.
[4] A Almog et al. *Front. Phys.* **16**, 2019.

## Contact Information

E-mail:
marzio.divece@imtlucca.it



## Method

In order to generate statistical models one, first, needs to find the corresponding $P(\mathbf{W})$. Our method to find $P(\mathbf{W})$ consists of three steps:

- constrained Shannon entropy maximization

$$S = -\sum_{\mathbf{W}} P(\mathbf{W}) \log P(\mathbf{W})$$

subject to the constraints defining the following hamiltonians:

$$\mathcal{H}_{\text{EGM}-1} = \alpha_0 L + \sum_i \sum_{j<i} (\beta_0 + \beta_{ij}) w_{ij}$$

where the constraints are the number of links and a tunable function of the weights;

$$\mathcal{H}_{\text{EGM}-2} = \sum_i \alpha_i k_i + \sum_i \sum_{j<i} (\beta_0 + \beta_{ij}) w_{ij}$$

where the constraints are the degrees and a tunable function of the weights;

Model TS adopts the topological step of the undirected binary Configuration Model (UBCM)

$$p_{ij} = \frac{e^{-\alpha_i - \alpha_j}}{1 + e^{-\alpha_i - \alpha_j}}$$

whereas model TSF adopts the topological step of the fitness model

$$p_{ij} = \frac{\delta \text{GDP}_i \text{GDP}_j}{1 + \delta \text{GDP}_i \text{GDP}_j}.$$

Both "dress" them with a geometric (exponential) distribution of the weights in the discrete (continuous) case.

- econometric transformation

$$\frac{y_{ij}}{1 - y_{ij}} = z_{ij} = \rho\, (GDP_i GDP_j)^{\beta} D_{ij}^{\gamma}$$

where $y_{ij} = e^{-\beta_{ij}}$ is the Lagrange parameter that drives the dyadic weighted term

- loglikelihood maximization

$$\mathcal{L} = \ln P(\mathbf{W}^* | \vec{\alpha}, \vec{\beta})$$

## Conclusions & Outlooks

### Conclusions

- We turned a set of Maximum Entropy Network models into Econometric models at different degrees of topological detail.
- We proved that also in the case of Econometric Network Model the degree-sequence input is an informative gain.
- We show that in terms of both Model Selection measures and higher-order statistics goodness-of-fit, EGM-3 outperforms all of the analyzed econometric counterparts.
- EGM is analytical and is applicable to any kind of weighted network, with continuous or discrete valued weights, directed or undirected, given any reasonable Gravity-like specification for the intensive margins.

### Outlooks

- Increase the complexity of the Gravity Specification using microfoundation.