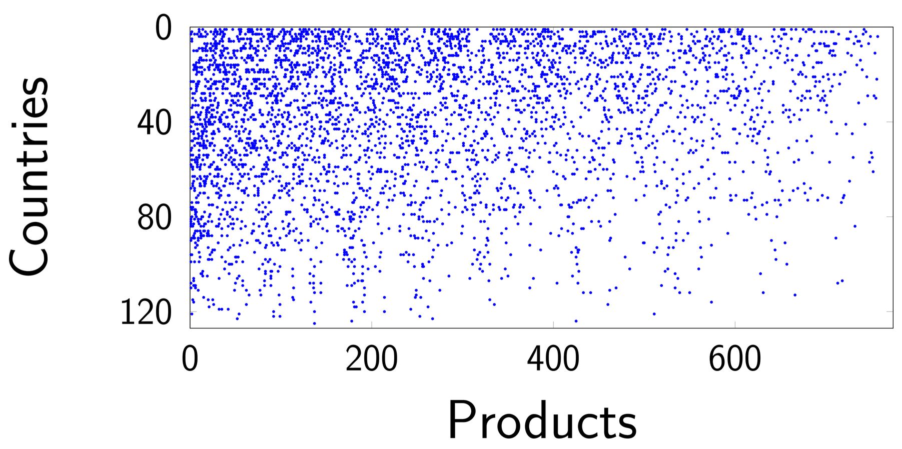
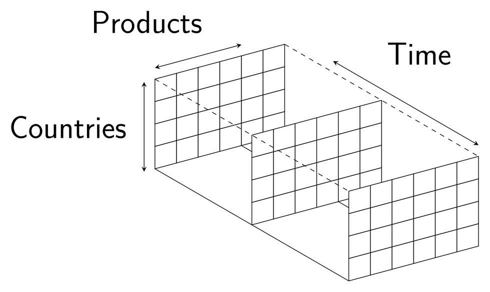
DYNAMIC POISSON FACTORIZATION MODEL FOR THE PRODUCTIVE STRUCTURE OF ECONOMIES

M. F. Pradier[1], V. Stojkoski[2], Z. Utkovski[2], L. Kocarev[2], F. Perez-Cruz[1]
[1]University Carlos III in Madrid,
[2]Macedonian Academy of Sciences and Arts
melanie@tsc.uc3m.es

PROBLEM STATEMENT

- **Aim**: What makes some countries wealthier than others?
  Hypothesis: Capabilities explain economic development of countries [3].
- **Contribution**: A BNP time-dependent Poisson factorization model to analyze international trade.
- **Key Idea**: Force sparsity in the features for interpretability, and account for temporal dynamics using a markov Indian buffet process.





STATIC POISSON FACTORIZATION MODEL

- Let $\mathbf{X} = \{0,1\}^{N\times D}$ be the thresholded revealed competitive advantage (RCA) matrix for N countries and D products:

$$\mathrm{RCA}_{nd} = \frac{E_{nd}/\sum_p E_{nd}}{\sum_n E_{nd}/\sum_{n,d} E_{nd}}$$

$$x_{nd} = \begin{cases} 1, & \text{if } \mathrm{RCA}_{nd} \geq 1 \\ 0, & \text{otherwise} \end{cases}$$

E_{nd} refers to the raw export size (in \$) for country n and product d.

- The generative model for the *Bernoulli Process Poisson Factor Analysis model* (BeP-PFA) is as follows:

$$x_{nd} \sim \mathrm{Poisson}(\boldsymbol{Z}_{n:}\, \boldsymbol{B}_{:d})$$
$$B_{kd} \sim \mathrm{Gamma}(\alpha_B, \frac{\mu_B}{\alpha_B})$$
$$\mathbf{Z} \sim \mathrm{IBP}(\alpha_Z)$$

- The IBP prior can be replaced by more flexible priors, such as the *Restricted IBP* or the *Three-parameter IBP*.

DYNAMIC POISSON FACTORIZATION MODEL

- We now assume T timestamps (years).
- We resort to a markov IBP to account for temporal dynamics [2].
- The generative model for the *dynamic Bernoulli Process Poisson Factor Analysis model* (dBeP-PFA) is as follows:

$$x_{nd}^{(t)} \sim \mathrm{Poisson}\left(\boldsymbol{Z}_{n:}{}^{(t)}\, \boldsymbol{B}_{:d}\right)$$
$$B_{kd} \sim \mathrm{Gamma}(\alpha_B, \frac{\mu_B}{\alpha_B})$$
$$a_k \sim \mathrm{Beta}(\frac{\alpha_Z}{K}, 1),$$
$$b_k \sim \mathrm{Beta}(\gamma, \delta),$$
$$z_{nk}^{(t)} | a_k, b_k \sim \mathrm{Bernoulli}\left(a_k^{1-z_{nk}^{(t-1)}}\, b_k^{z_{nk}^{(t-1)}}\right)$$

where $z_{nk}^{(0)} = 1, \forall n, k$. The transition matrix Q_k for feature k is given by:

$$Q_k = \begin{pmatrix} 1-a_k & a_k \\ 1-b_k & b_k \end{pmatrix}$$

INFERENCE

- MCMC approach, e.g., Gibbs sampler + slice sampler for the IBP
- K Poisson-distributed auxiliary random variables, i.e., $x_{nd}^{(t)} = \sum_{k=1}^K r_{nd,k}^{(t)}$
- Forward Filtering Backward Sampling (FFBS) to approximate $p(\mathbf{Z}|\mathbf{X}, \mathbf{B})$

$$p(\mathbf{X}_{n:}^{(1:t)}, z_{nk}^{(t)} | -) = p(\mathbf{X}_{n:}^{(t)} | z_{nk}^{(t)}, -) \sum_{z_{nk}^{(t-1)}} p(\mathbf{X}_{n:}^{(1:t-1)}, z_{nk}^{(t-1)} | -)\, p(z_{nk}^{(t)} | z_{nk}^{(t-1)})$$

- Forward step: compute $p(z_{nk}^{(t)} | \mathbf{X}_{n:}^{(1:t)}, \mathbf{Z}_{n,\neg k}^{(t)}, \mathbf{B})$
- Backward step: sample from $p(z_{nk}^{(t)} | z_{nk}^{(t+1)}, \mathbf{X}_{n:}^{(1:t)}, \mathbf{Z}_{n,\neg k}^{(t)}, \mathbf{B})$

RESULTS

We compare against:

- Poisson-Gamma Dynamical Systems (PGDS) [4]
- Thinned Gamma Process Poisson Factor Analysis (tGaP-PFA) [1]

Metric	PGDS	tGaP-PFA	BeP-PFA	dBeP-PFA
Test LLH	-0.648 ± 0.002	-0.356 ± 0.001	$\mathbf{-0.324 \pm 0.003}$	-0.350 ± 0.005
Coherence	-	-469.11 ± 9.562	-506.29 ± 13.470	$\mathbf{-403.70 \pm 31.725}$

Table 1 : Model Comparison in terms of test log-likelihood and topic coherence.

Id	Top-3 products with highest weights
F0	(bias) crude petroleum, crustaceans, cereals
F1	light fixtures, locksmith hardw., misc. ceramic ornaments
F2	inorganic esters, chemical products, nitrogen compound
F3	iron sheets, iron wire, thin iron sheets
F4	misc. elect. machinery, typewriters, misc. office equipment
F5	soaps, confectionary sugar, baked goods
F6	bovine – equine entrails, bovine meat, misc. prepared meats
F7	knit clothing accessories, linens, leather accessor.
F8	glazes, textiles fabrics for machinery, mineral wool
F9	misc. vegetables, grapes – raisins, misc. fruit
F10	inorganic bases, nitrogenous fertilizers, lubricating petrol. oils
F11	imitation jewellery, embroidery, synth. precious stones
F12	coffee, non-coniferous worked wood, cane sugar
F13	copper ores, chemical wood pulp, misc. non-ferrous ores
F14	pepper, vegetable planting materials, natural rubber
F15	raw cotton, cotton linters, green groundnuts

Table 2 : List of inferred latent features.

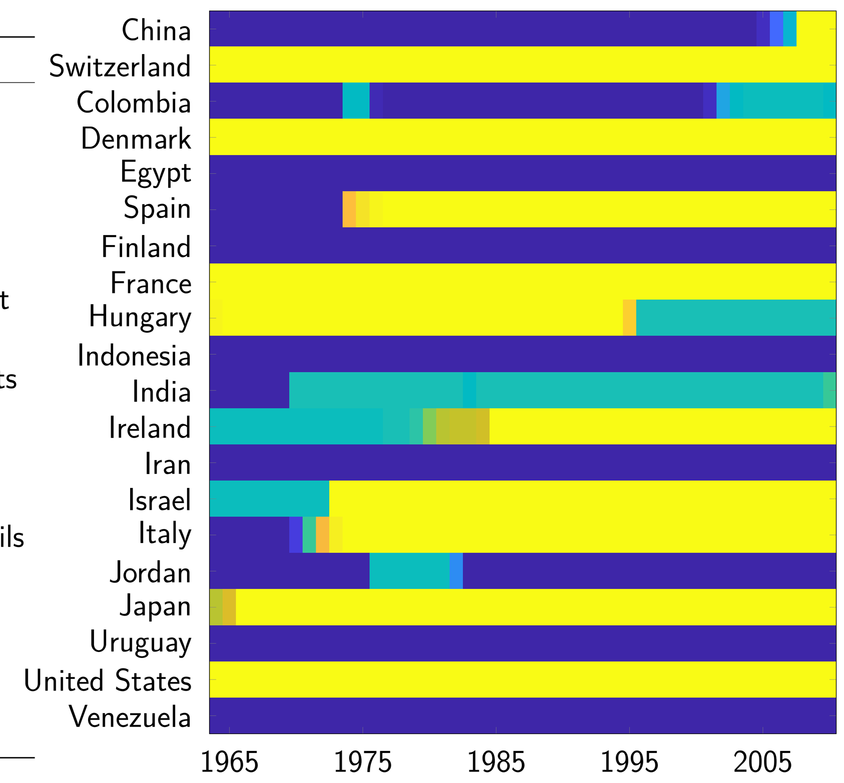


Figure 1: $\boldsymbol{Z}_{:k}^{(:)}$: Feature activation for $k = 2$.

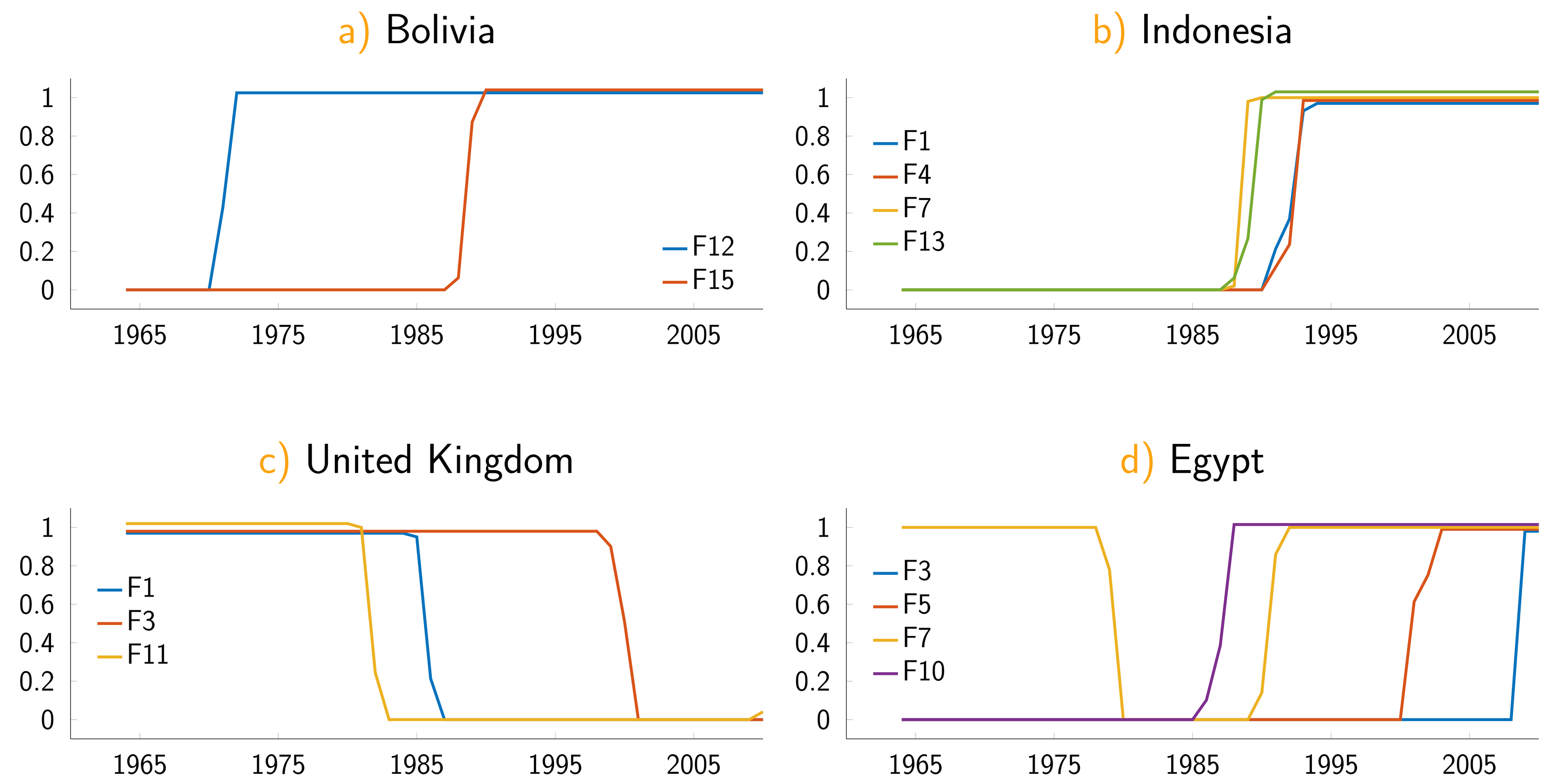


Figure 2: $\boldsymbol{Z}_{n:}^{(:)}$: Features activation for specific countries. Always active features omitted, e.g., Bolivia: F13; Indonesia: F12, F14; United Kingdom: F2, F4, F5, F8, F10; Egypt: F9, F15.

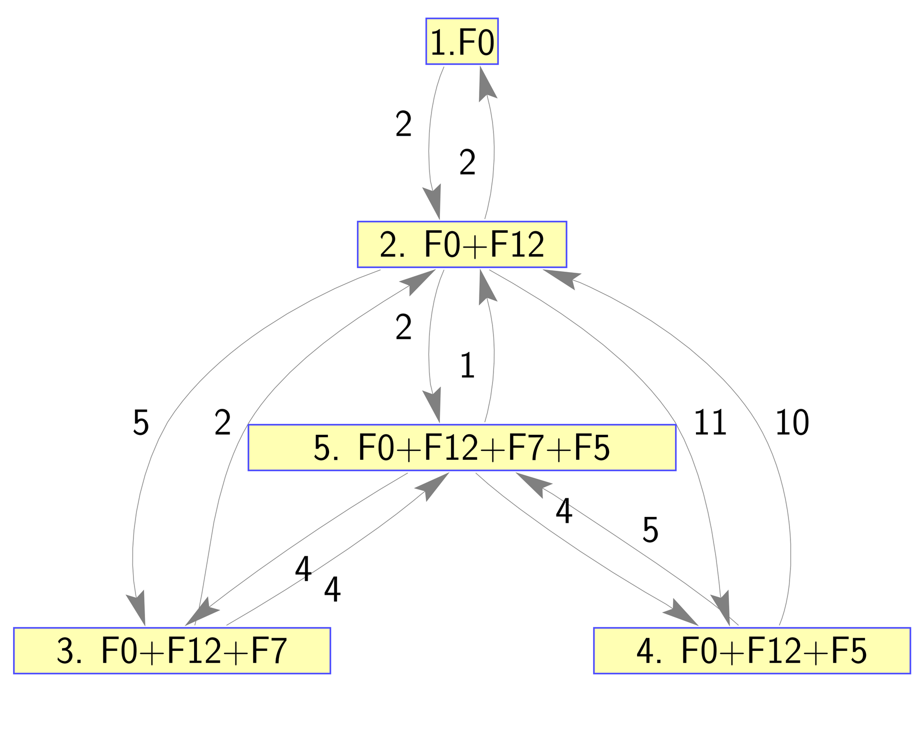


Figure 3: Finite state machine of latent features.

Transitions	Country	Years
1 → 2	Congo	1996-1997
2 → 1	Angola	1976-1977
2 → 3	Colombia	1977-1978
2 → 4	Cote D'Ivoire	1998-1999
2 → 4	Cameroon	1970-1971
2 → 4	Ecuador	1992-1993
2 → 5	El Salvador	1985-1986
3 → 2	Cambodia	1996-1997
3 → 5	Costa Rica	1998-1999
4 → 2	Cote D'Ivoire	1983-1984
4 → 2	Cameroon	1975-1976
4 → 2	Honduras	1997-1998
4 → 2	Nicaragua	1982-1983
4 → 5	Costa Rica	1986-1987
5 → 2	El Salvador	1982-1983
5 → 3	Costa Rica	1997-1998
5 → 4	Costa Rica	1999-2000

Table 3 : Transition examples.

CONCLUSIONS

1. **Interpretable** BNP model for temporal high-dim count data (sparse, non-negative features).
2. **Markovian dynamics** for features activation.
3. Analysis of productive structure of **world economies**.

WORK IN PROGRESS

- **Mixing improvement**, better coverage of the posterior:
  - Split and merge moves.
  - Particle Gibbs with Ancestor Sampling.
- **Dictionary variation over time** via Gamma-Poisson auto-regressive chains.

REFERENCES

1. Nicholas J. Foti, Joseph D. Futoma, Daniel N. Rockmore, and Sinead Williamson.
A unifying representation for a class of dependent random measures. 2012.
2. Jurgen V. Gael, Yee W. Teh, and Zoubin Ghahramani.
The Infinite Factorial Hidden Markov Model. 2009.
3. César A. Hidalgo and Ricardo Hausmann.
The building blocks of economic complexity. 2009.
4. Aaron Schein, Hanna Wallach, and Mingyuan Zhou.
Poisson-Gamma dynamical systems. 2016.

SIGNAL PROCESSING GROUP
DEPARTMENT OF SIGNAL THEORY AND COMMUNICATIONS
UNIVERSIDAD CARLOS III DE MADRID