

---

Πατησίων 76, Αθήνα 104 34 Greece. Τηλ.: (+30) 2108203112-3 / [www.aueb.gr](http://www.aueb.gr)  
Ιωάννης Ντζουφρας, Τηλ. (+30) 2108203968, E-mail: [ntzoufras@aubg.gr](mailto:ntzoufras@aubg.gr)  
[www.stat-athens.aueb.gr/~jbn](http://www.stat-athens.aueb.gr/~jbn)

---

## ΔΕΛΤΙΟ ΤΥΠΟΥ

### International HERMES Ph.D. Workshop 2023: “Data Science in Business”



Το Τμήμα Στατιστικής του Οικονομικού Πανεπιστημίου Αθηνών διοργάνωσε, στις 7 και 8 Ιουνίου 2023, με μεγάλη επιτυχία, το International HERMES Ph.D. Workshop 2023: “Data Science in Business”, που έλαβε χώρα στις εγκαταστάσεις του Οικονομικού Πανεπιστημίου Αθηνών.

Οι συμμετέχοντες του Workshop προέρχονταν από πληθώρα Ευρωπαϊκών Πανεπιστημίων και, πιο συγκεκριμένα, από τα:

- [Academia de Studii Economice din București](#) (Ρουμανία)
- [Department of Statistics, Athens University of Economics and Business](#) (Ελλάδα)
- [International Hellenic University](#) (Ελλάδα)
- [Leopold-Franzens Universität Innsbruck](#) (Αυστρία)
- [Maynooth University](#) (Ιρλανδία)
- [Technische Universität Dresden](#) (Γερμανία)
- [Università Ca' Foscari Venezia](#) (Ιταλία)
- [Universidad de Alcalá](#) (Ισπανία)
- [Università di Pavia](#) (Ιταλία)
- [University of Economics in Bratislava](#) (Σλοβακία) και
- [University of Lausanne](#) (Ελβετία).

Το Workshop περιλάμβανε 12 Παρουσιάσεις (“*Oral Presentations*”) και 10 ανακοινώσεις Poster (“*Poster Presentations*”), οι οποίες συνοδεύονταν από άτυπη συζήτηση και ανταλλαγή ιδεών μεταξύ των συμμετεχόντων, καθιστώντας αυτήν την εκδήλωση την ιδανική ευκαιρία για εύστοχη, πολυπολιτισμική ερευνητική ανταλλαγή γνώσεων, αλλά και δικτύωση, μεταξύ Υποψήφιων Διδακτόρων διαφόρων γνωστικών αντικειμένων.

Την πρώτη ημέρα (7 Ιουνίου 2023), οι συμμετέχοντες συγκεντρώθηκαν στο Αμφιθέατρο (Επίπεδο -1), του Νέου Κτιρίου του ΟΠΑ ([Troias 2 and Spetson](#)), εκεί όπου τους υποδέχτηκαν, ο Αντιπρύτανης του ΟΠΑ, Καθηγητής Βασίλης Παπαδάκης, ο Πρόεδρος του Τμήματος Στατιστικής του ΟΠΑ, Καθηγητής Ιωάννης Ντζούφρας, αλλά και η Πρόεδρος του HERMES, Elke Kitzelmann (Associate Dean of Studies, International Economic & Business Studies, University of Innsbruck):





Ακολούθησε η Κεντρική Ομιλία (“*Keynote Speech*”) του Άγγελου Αλεξόπουλου (ΟΠΑ), με θέμα: “*A Machine Learning and Network approach to Value Added Tax Fraud Detection*”:

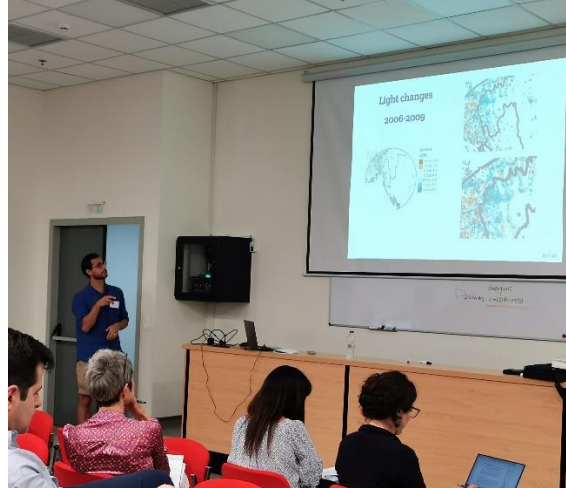


Μετά το πέρας της ομιλίας, οι συμμετέχοντες απόλαυσαν ένα μικρό διάλλειμα, που τους έδωσε την ευκαιρία ν’ αρχίσουν να γνωρίζονται καλύτερα μεταξύ τους:



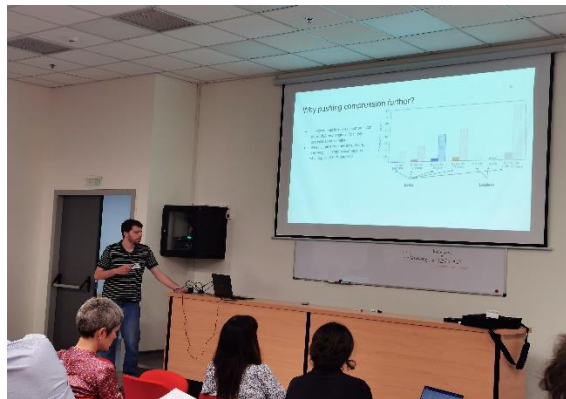
Η πρώτη Συνεδρία του Workshop, με προεδρεύων τον Bernhard Schipp (*T.U. Dresden*), αποτελούνταν από τις κάτωθι ομιλίες:

1. Juan Felipe Santos Marquez (*Technische Universität of Dresden*): [Tri-border Areas and the Location of Economic Activity in Open Economies](#):



2. Miruna Proscanu and Cosmin Proscanu (*Bucharest University of Economic Studies*): [Electricity consumption forecasts for Romania. Deep Learning versus Time Series models](#) και

3. Xenophon Kitsios (*AUEB*): [A Time Series Compression Technique](#):



Τις Προφορικές Παρουσιάσεις (“*Oral Presentations*”) διαδέχθηκαν οι ανακοινώσεις Poster (“*Poster Presentations*”), συνοδευόμενες κι από ένα διάλειμμα για γεύμα, στο Αμφιθέατρο Αντωνιάδου, του Κεντρικού Κτιρίου του ΟΠΑ ([Patission 76](#)):







## 2. Parametric Survival Modeling of Soccer Data (Ilias Leriou, AUEB):

### PARAMETRIC SURVIVAL MODELING OF SOCCER DATA.

Ilias Leriou\*, Ioannis Ntzoufras\* and Dimitris Karfis\*  
\*Athens University of Economics and Business

**Our goal (figuratively!)**

- Find a plausible parametric distribution for Bayesian survival modeling.
- Explore bivariate goal arrival time modeling.
- Predictive league reconstruction.

**What we know so far**

- Del Corral (2008)
  - Analysis of first substitution time and their determinants in Spanish league for season 2004-5.
- Novo (2013)
  - Cox model for 1st & 2nd goal. 760 Premier League games (2 seasons, 2008-2010).
- Egidi (2018)
  - Use of dynamic time dependent parameters used to capture the performance of the teams.
- Tsokos (2019)
  - Machine learning tools for modeling soccer events.
- Narayanan (2021)
  - Model football association event times using Hawkes processes.

**Data Layout**

Let  $t_{1i}$  and  $t_{2i}$  be the event gap times for team 1 and team 2 respectively with  $i = 1, 2, \dots, n$  and  $m = 1, 2, \dots, M$  the game indicator.

Game	Home	Away	Home	Away
1	Z	NA	0	2
1	SS	NA	0	80
1	NA	9	9	0
1	NA	NA	4	4

**Bayesian Model Discrimination**

The model's structure is presented below:

$T_{ij} \sim Weibull(\gamma, \lambda_j), j = 1, 2, i = 1, 2, \dots, n$

with  $\begin{cases} E(T_{1i}) = \mu + \alpha_{2Z} + \theta_{2Z} + d_{2Z} \\ \log E(T_{2i}) = \mu + \alpha_{2Z} + d_{2Z} \end{cases}$

where

$E(T_{ij}) = \lambda_j^{-1} \Gamma(1 + 1/\gamma)$

with  $i = 1, 2, \dots, n, j = 1, 2$

Weakly informative priors to parameters are as follows:

$\alpha_{2Z}, d_{2Z}, \mu, \theta_{2Z} \sim Normal(0, 10^{-3})$

while the following weakly informative Gamma prior was assigned to the positive parameter  $\gamma$ :

$\gamma \sim Gamma(10^{-3}, 10^{-3})$

STZ constrains for attacking and defensive parameters to allow for comparisons of the abilities of each team with the overall level of the fixed effects:

$\sum_{k=1}^K \alpha_k = 0, \sum_{k=1}^K d_k = 0$

**Inference**

	Mean	Median	sd	2.5%	97.5%
$\mu$	4.800	4.807	0.058	4.655	4.892
$\gamma$	1.117	1.116	0.027	1.066	1.173
home	-0.258	-0.262	0.058	-0.365	-0.134

**How do we do?**

Team (EPL 2018/2019)	Predicted (actual) ranking*
Manchester City	1 (1)
Liverpool	2 (2)
Tottenham	3 (4)
Chelsea	4 (3)
Arsenal	5 (5)
Manchester United	6 (6)
Everton	7 (9)
Leicester	8 (7)
Wolverhampton Wanderer	9 (8)
Crystal Palace	10 (10)
West Ham	11 (13)
Watford	12 (11)
Newcastle United	13 (12)
Bournemouth	14 (14)
Southampton	15 (16)
Burnley	16 (15)
Brighton	17 (17)
Cardiff	18 (18)
Fulham	19 (16)
Huddersfield	20 (20)

**Acknowledgements**

The research work was supported by the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under the HFRI PhD Fellowship grant 186756/20/01/11/2017.

## 3. Density Estimates for the S-Laplacian And Applications (Michael Nikolouzos, AUEB):

### DENSITY ESTIMATES FOR THE S-LAPLACIAN AND APPLICATIONS

M. Nikolouzos<sup>1\*</sup>, A. N. Yannacopoulos<sup>1\*</sup>  
<sup>1</sup>Athens University of Economics and Business, <sup>\*</sup>Stochastic Modeling and Applications Laboratory

**Intuition**

What is the connection between Levy processes and minimal solutions of a fractional Poisson system of equations? Let  $N_t$  a (2S-stable) Levy process and  $u$  a minimal solution of the system  $(-\Delta)^s u + W(x)u = 0$ , for  $u: \mathbb{R}^m \rightarrow \mathbb{R}^m, m \geq 1, 0 < s < 1$  and  $W$  the potential. The fractional Laplacian is defined as

$$(-\Delta)^s u(x) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon^{2s}} \int_{\mathbb{R}^m} (u(x) - u(y)) dy$$

Following the work of [1] and [2], we are proving a density theorem for the above system, which will provide a lower bound for the energy functional (defined below). The importance of the theorem is that provides a pointwise estimation for the vector function  $u$ .

**Main Theorem**

We consider minimizers of the nonlocal energy functional  $J(u, \Omega) = \frac{1}{2} \int_{\Omega \times \Omega} \frac{|u(x) - u(y)|^2}{|x - y|^{2s}} dx dy + \int_{\Omega} W(u(x)) dx, s \in (0, 1)$ , among all functions  $u: \Omega \subset \mathbb{R}^m \rightarrow \mathbb{R}^m$ , such that  $u \in W^{s,2}(\Omega; \mathbb{R}^m)$ , with  $u = g$  (given) on  $\mathbb{R}^m \setminus \Omega$ , for a given open bounded set  $\Omega \subset \mathbb{R}^m$ , where  $|\cdot|$  denotes the Euclidean distance (in  $\mathbb{R}^m$  or  $\mathbb{R}^n$ ) and  $W: \mathbb{R}^m \rightarrow \mathbb{R}$  is a  $C^1$  class (Fig.1), positive potential, for  $\alpha \in (0, 2]$ , with a zero at  $\alpha \in \mathbb{R}^m$ , satisfying the hypothesis  $\exists r_0 > 0, \forall \xi \in \mathbb{R}^m, |\xi| = 1, (0, r_0] \ni r \rightarrow W(\alpha + r\xi)$  is non decreasing with  $W(\alpha + r_0\xi) > 0$ .

Furthermore the term minimizer means that for any  $v$  in the same class as  $u$  with  $u = v$  in  $\Omega^c$ , the inequality  $J(u, \Omega) \leq J(v, \Omega)$  holds.

**Theorem:** Let  $s \in (0, 1)$  and assume that the potential satisfies the above assumptions,  $\Omega$  is open and  $u: \Omega \subset \mathbb{R}^m \rightarrow \mathbb{R}^m$  is minimal. Then, for any  $\mu_0 > 0$ , and any  $\lambda \in (0, \mu_0]$  where  $d_\lambda = \inf_{\Omega} \{ |u - \lambda| : u \neq \alpha, W(x) = 0 \} > 0$ , the condition

$$|B_{\lambda}(x_0) \cap \{ |u - \alpha| > \lambda \}| \geq \mu_0$$

implies that

$$|B_{\lambda}(x_0) \cap \{ |u - \alpha| > \lambda \}| \geq Cr^{\alpha}, \text{ for } r \geq r_0$$

as long as  $B_{\lambda}(x_0) \subset \Omega$  where  $C = C(W, \mu_0, \lambda, r_0, M)$ .

**Progress**

So far, we have managed to prove the main theorem for the cases:

- $\bullet 0 < s < 1$  and  $\alpha = 2$
- $\bullet 0 < s < 1$  and  $1 < \alpha < 2$

We, also, have proven an upper bound for the energy of the minimizer:

$$J(u, B_R) \leq \begin{cases} CR^{2s-1} & \text{if } s \in (1/2, 1) \\ CR^{2s-1} \ln R & \text{if } s = 1/2 \\ CR^{2s-2} & \text{if } s \in (0, 1/2) \end{cases}$$

for an appropriate constant  $C$  independent of  $R$ .

In addition, a first application of the main theorem is the derivation of a pointwise estimate for the minimal solutions of the system for the case  $\alpha = 2$  and  $s \in (0, 1)$ . Under the assumptions of the main theorem and for a given  $I$ , there exists  $R(I)$  (depending only on  $W$  and  $M$ ) such that:  $B_{R(I)}(x_0) \subset \Omega$  implies  $|u(x_0) - \alpha| < I$ .

**What's next**

Currently, we are working on the rest cases of the main theorem:

- $\bullet 0 < s < 1$  and  $0 < \alpha < 1$

The above case it has been proven to be difficult and currently, we are redesigning the entire proof using a new family of test functions. The functions of the family  $f_\lambda(x) = -(1 + |x|^2)^{-\lambda/2}$ , for suitable  $\lambda$ , are s-subharmonic for  $s \in \mathbb{R}^m$  and enjoys many more properties that are suitable for our case.

Moreover, the pointwise estimate above will conclude on a Liouville type theorem for minimizers. We are investigating more aspects of the theorem and, in general, of the system:

- $\bullet$  connections (minimal solutions that connect the minima of the potential).
- $\bullet$  stratification-hierarchical structure of the system under symmetry hypotheses.

**References**

- [1] N. D. Alkaios, G. Fusco, and P. Smyrnelis. *Elliptic systems of phase transition type*. PNLDE 91, Birkhäuser (Green Series), 2018.
- [2] O. Savin and E. Valdinoci. Density estimates for a variational model driven by the gaillardot norm. *Journal de Mathématiques Pures et Appliquées*, 101(1):1–26, 2014.



#### 4. Deep Learning Models for Probability of Default (Kyriakos Georgiou, AUEB):

DEEP LEARNING MODELS FOR PROBABILITY OF DEFAULT

Kyriakos Georgiou<sup>1</sup> and Athanasios N. Giannakopoulos<sup>1</sup>

<sup>1</sup>AUEB - Department of Statistics

**Motivation**

The International Financial Reporting Standards (IFRS) 9 have amplified the need for rigorous mathematical methods, able to quantify, assess and optimize credit risk. We consider the problem of estimating Probabilities of Default (PD) of asset stochastic processes by using Deep Neural Networks to train models that predict these values, motivated by recent developments in Deep Learning as tools for solving PDEs.

**The asset process models**

We consider a "generalized" Lévy-driven stochastic process, under which the asset value process is defined by the triple  $(G_t, R_t, Y_t)_{t \geq 0}$ , capturing both the switching and volatility processes, and is given by:

$$dG_u = k(R_u)(\theta(R_u) - G_u)dt + \sigma(R_u)\sqrt{Y_u}dB_u + \int_{\mathbb{R}} zN(dz, dt)$$

$$dY_u = \kappa(\mu - Y_u)dt + \xi\sqrt{Y_u}dW_t$$

with  $G_0 = x, R_0 = \rho$  and  $Y_0 = y$ .

**PIDEs for the PD function**

We define the PD as a function of the initial values and the time until maturity:

$$\Psi(x, \rho, y, u) = \mathbb{P}\left(\inf_{0 \leq t \leq u} G_s \leq 0 \mid G_0 = x, R_0 = \rho, Y_0 = y\right)$$

and survival probability  $\Phi(x, \rho, y, u)$ . We have shown that the  $\Phi$  is the solution of the PIDE:

$$\frac{\partial \Phi}{\partial u} = k(\rho)\theta(\rho) - x\frac{\partial \Phi}{\partial x} + \kappa(\mu - y)\frac{\partial \Phi}{\partial y} + \frac{1}{2}\sigma^2\rho^2\frac{\partial^2 \Phi}{\partial x^2} + \frac{1}{2}\xi^2 y\frac{\partial^2 \Phi}{\partial y^2} + \sum_{j \neq \rho} a_{j\rho}(\Phi(x, j, y, u) - \Phi(x, \rho, y, u)) + \int_{\mathbb{R}} (\Phi(x+z, \rho, y, u) - \Phi(x, \rho, y, u))\nu(dz)$$

with initial and boundary conditions:

$$\Phi(x, \rho, y, 0) = \mathbf{1}_{x > 0}, \Phi(0, \rho, y, u) = 0, \Phi(x, \rho, y, u) \rightarrow 1, \text{ as } x \rightarrow \infty, \frac{\partial \Phi}{\partial y}(x, y, u) = 0, \text{ as } y \rightarrow \infty$$

Standard numerical solutions suffer from the "curse of dimensionality":

**Training a Neural Network**

We can use the Feynman-Kac formula to create a loss function using an appropriate payoff function. Consider the random variable  $Y^{(x, \rho, y)} = h(T, G_T^{(x, \rho, y)}, \mathbf{1}(\inf_{0 \leq t \leq T} G_t^{(x, \rho, y)} < 0))$ . Then:

$$\Phi(x, t, \rho, y) = \mathbb{E}[h(T, G_T^{(x, \rho, y)}) | \mathcal{F}_t] = \mathbb{E}[h(T, X_T) | X_t = x]$$

where the equality is a result of the Markov property, and therefore we can write:

$$\Phi(x, t, \rho, y) = \mathbb{E}[Y^{(x, \rho, y)} | \mathcal{F}_t]$$

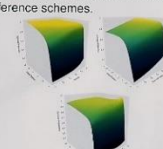
Hence,  $\Phi$  is the solution to the minimization problem:

$$\min_{\theta} \mathbb{E}[|Y^{(x, \rho, y)} - \Phi(x, \rho, y)|]$$

To train a DNN model we can use the estimator of the expectation above as the loss function:

**Neural Network PD models**

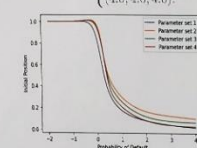
The resulting survival functions under the generalized, regime switching and stochastic volatility models are displayed (from left to right) below. These closely match solutions obtained using standard Finited Difference schemes.



**DNN for a family of asset models**

Using the DNNs we can therefore solve the "curse of dimensionality" and extend their usage. In practice it may be useful to consider an asset value process with parameter vector  $\Theta = \{\theta_1, \dots, \theta_n\}$ , which follows an appropriate multidimensional distribution function  $\Theta \sim F$ . Let the one dimensional asset value model (i.e.,  $G_u$  with a single regime and constant volatility) where the

stochastic coefficients  $\Theta = (k, \theta, \sigma)$  are each randomly sampled from  $Uniform(0.0, 5.0)$ . We can train a DNN model that has a four-dimensional input layer  $(x, \rho, k, \theta, \sigma)$ . Below we display the resulting PD functions predicted by the DNN for four parameter sets:

$$(k, \theta, \sigma) = \begin{cases} (0.5, 0.5, 0.5), \\ (2.0, 2.0, 2.0), \\ (3.0, 3.0, 3.0), \\ (4.0, 4.0, 4.0). \end{cases}$$


This displays the important benefit these Machine Learning models can offer, as we can generalize the input layer to account for multiple parameters of the models.

**Future work**

Research pertaining to DNNs still has many open questions:

- Large errors can occur near initial and boundary conditions.
- Activation functions that use simplified versions of the models.
- Comparison with "Physics Informed Neural Networks".

#### 5. Model-Based Clustering for Dynamic Count-Valued Social Networks (Angelos Kekempanos, AUEB):

MODEL-BASED CLUSTERING FOR DYNAMIC COUNT-VALUED SOCIAL NETWORKS

Kekempanos Angelos

Athens University of Economics and Business

**introduction**

The Dynamic Count-Valued Social Networks (DCVSN) are social networks that:

- Consist of  $N$  nodes.
- Measure the pairwise relationships between the nodes using counts (e.g. mails, calls, number of events etc.).
- The number of events between the nodes change over time (dynamic evolution).

Here, I introduce two model-based algorithms that detect communities-clusters in a DCVSN.

**Models' Definition**

Suppose that a DCVSN can be described by a count-valued adjacency cube  $Y_{ij}^{(t)}$  for  $i, j = 1, \dots, N$  and  $t = 1, \dots, T$ .

**Dynamic Latent Space GLM**

Assume that:

$$Y_{ij}^{(t)} \sim \text{Poisson}(\lambda_{ij}^{(t)})$$

The rate parameter  $\lambda_{ij}^{(t)}$  is modelled as:

$$\log(\lambda_{ij}^{(t)}) = \beta X_{ij} + \gamma^{(t)} \mathbf{1}_{\{Y_{ij}^{(t-1)} > 0\}} + \delta^{(t)} \mathbf{1}_{\{Y_{ij}^{(t-1)} = 0\}} - \|W_i - W_j\|$$

- $\gamma^{(t)}$  represents the total increasing tendency of the count-interactions in the network over time.

**Estimation**

**Bayesian Inference**

We estimate the parameters of the models using a Bayesian inference approach.

- The posterior distributions were calculated after assuming the proper prior distribution.
- A combination of Metropolis-Hastings and Gibbs samplers were used for the computational estimation of the models' parameters.

**Label Switching**

We solve the label switching problem using the Equivalence Classes Representatives (ECR) algorithm.

**Number of Clusters**

We choose the models with the maximum values of BIC approximations.

For the Dynamic Latent Space GLM

$$BIC_{DLSGLM} = BIC_{DGLM} + BIC_{FMG}$$



For the Poisson Autoregressive LSM

$$BIC_{PALSM} = BIC_{INAR} + BIC_{FMG}$$

**Implementation**

The Toronto bikeshare network was used for the application of the models to real world data.

The edges describe the total number of shared bikes among the 137 busiest stations for each month from 01-2020 to 10-2021. Both models revealed 4 quite reasonable clusters-communities.


**References**

- Mark S. Handcock, Adrian E. Raftery, Jeremy M. Tantrum (2007). "Model-Based Clustering for Social Networks"
- Niall Friel, Riccardo Rastelli, Jason Wynn and Adrian E. Raftery (2016). "Interlocking directorates in Irish companies using a latent space model for bipartite networks"
- Pappadamos, P. (2014). Handling the label switching problem in latent class models via the ECR algorithm.

6. Bayesian analysis of diffusion-driven multi-type epidemic models with application to COVID-19 (Lampros Bouranis, AUEB):

# Bayesian analysis of diffusion-driven multi-type epidemic models with application to COVID-19

Lampros Bouranis<sup>\*1,1</sup>, Nikolaos Demiris<sup>1</sup>, Konstantinos Kalogeropoulos<sup>2</sup> and Ioannis Ntzoufras<sup>1</sup>  
<sup>1</sup>Department of Statistics, AUEB, Athens, Greece <sup>2</sup>Department of Statistics, LSE, London, United Kingdom  
 \*bouranis@aueb.gr



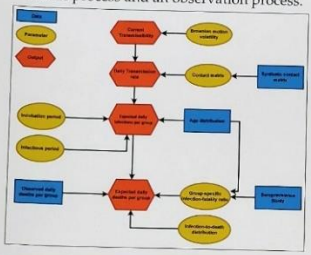
---

## Objectives

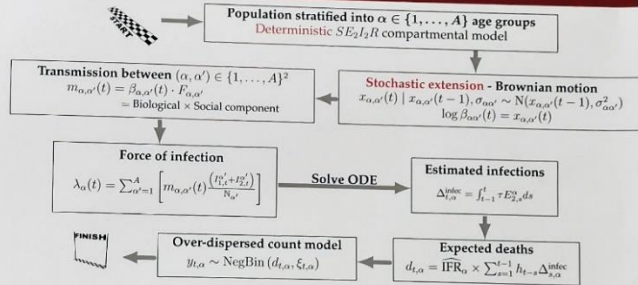
We consider a flexible Bayesian evidence synthesis approach to model the age-specific transmission dynamics of COVID-19 based on daily mortality counts. The temporal evolution of transmission rates in populations containing multiple types of individual is reconstructed via an appropriate dimension-reduction formulation driven by independent diffusion processes. A suitably tailored compartmental model is used to learn the latent counts of infection, accounting for fluctuations in transmission influenced by public health interventions and changes in human behaviour.

### Bayesian evidence synthesis

- Estimation of hidden characteristics of the disease like the latent number of infections.
- Separate modeling process into a latent epidemic process and an observation process.



### Modeling framework



---

## Parameter estimation

- The model is viewed as a hypo-elliptic diffusion - Intractable! Solution of non-linear system of ODEs approximated with the Trapezoidal rule (forward simulation).

## Key notes:

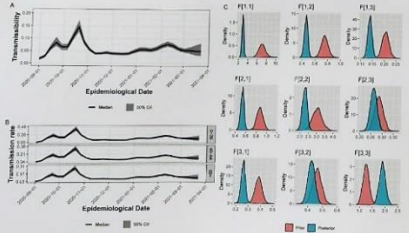
- Let  $y_{t,\alpha}$  be the number of observed deaths on day  $t = 1, \dots, T$  in age group  $\alpha \in \{1, \dots, A\}$ . A given infection may lead to observation events (i.e. deaths) in the future.
- The latent epidemic process is expressed by ordinary differential equations (ODEs).
- We target the transmission rate matrix process  $m_{\alpha,\alpha'}(t)$  whose dimension increases quadratically with  $A$ .
- Facilitates model determination at a latent level, performed by appropriate model expansion.

---

## Application

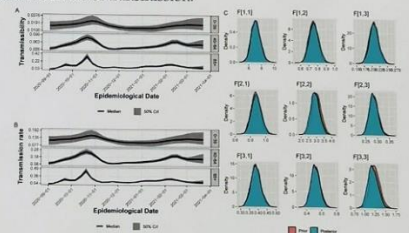
### 1. The COVID-19 pandemic

- Greece: August 2020 – March 2021.
- Model  $m_{\alpha,\alpha'}(t) = \beta_t \cdot F_{\alpha,\alpha'}$  not flexible enough to accommodate for age-specific trends in SARS-CoV-2 transmission.
- Gain essential identifiability from model expansion.



### 2. Model expansion

- Allow for age-specific time-varying transmissibilities, such that  $m_{\alpha,\alpha'}(t) = \beta_t^{\alpha\alpha'} \cdot F_{\alpha,\alpha'} = \beta_t^\alpha \cdot F_{\alpha,\alpha'}$ , under the assumption  $\beta_t^{\alpha\alpha'} = \beta_t^\alpha$ ,  $\alpha \neq \alpha'$ , for reasons of parsimony.
- Model (1) with independent BMs enables reconstruction of the age-specific drivers of transmission.



---

## Discussion

- Model (1) was validated using the estimated age-specific numbers of cumulative infections in England from the REACT-2 seroprevalence survey.
- Future work: Age-specific forecasting of deaths; Model expansion - Exchangeable Brownian Motions.
- arXiv preprint: <https://arxiv.org/abs/2211.15229>
- R package: <https://CRAN.R-project.org/package=Bernadette>

---

## Acknowledgements

Lampros Bouranis' research is supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101027218.

---

## References

- Flaxman, S. et al (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820):257–261.
- Knock, E. et al (2021). Key epidemiological drivers and impact of interventions in the 2020 SARS-CoV-2 epidemic in England. *Science Translational Medicine*, 13(602):eabg4262.



## 7. Bayesian Spatio-Temporal Epidemic Models (Petros Barmounakis, AUEB):

BAYESIAN SPATIO-TEMPORAL EPIDEMIC MODELS  
P. Barmounakis and N. Demiris

**Study Objectives**

The main contribution of this work is the introduction of different stochastic differential models embedded in zero-inflated epidemic models and their evaluation on livestock epidemic data from Evros, Greece.

**Zero-inflation model**

We use a zero-inflation model to account for the excess zeros and to examine the parameters/covariates that contribute to a disease-free environment.

$$y_i \sim g(y_i | \lambda_i, p_i)$$

$$g(y_i | \lambda_i, p_i) = p_i I_{(y_i=0)} + (1-p_i) f(y_i | \lambda_i)$$

$$\lambda_i = \int_{t_i-1}^{t_i} \exp(\lambda_s) ds, \quad i = 1, \dots, T$$

$$\mu_i = X_i \beta + K(d_i, \Theta_k)$$

$$p_i = X_i \beta^2 + K^2(d_i, \Theta_k)$$

where  $f(\cdot)$  is the probability mass function of the Poisson distribution and  $\lambda_i$  is the rate.  $X_i$  is the design matrix containing information about previously infected villages and meteorological data. The terms  $K(d_i, \Theta_k)$  and  $K^2(d_i, \Theta_k)$  are infection kernels, where  $d_i = \{d_{ij} : k \in S_i, l \in I_{i-j}\}$ , is the set of all Euclidean distances between previously infected ( $I_{i-j}$ ) and uninfected ( $S_i$ ) farms at time  $i$  within the typical infectious time of the disease.

**Ornstein-Uhlenbeck model for the log-rate**

$$\lambda_i = \int_{t_i-1}^{t_i} \exp(\lambda_s) ds, \quad i = 1, \dots, T$$

$$d\lambda_t = \phi(\lambda_t - \mu_t) + \sigma dW_t$$

$$\lambda_{t+1} | \lambda_t \sim N(\mu_{t+1} + (\lambda_t - \mu_t) e^{-\phi}, \frac{\sigma^2}{2\phi} (1 - e^{-2\phi}))$$

**OU with Student's t marginals**

A l.v.  $X \sim T(\nu, \tilde{\mu}, \tilde{\delta})$  with  $X \stackrel{D}{=} \tilde{\mu} + \tilde{\sigma} \epsilon$  where  $\epsilon \sim N(0, 1)$  and  $\tilde{\sigma}^2 \sim \text{Inv}(\frac{\nu}{2}, \frac{1}{2} \tilde{\delta}^2)$  follows the Student's t-distribution.

$$\lambda_{t+1} | \lambda_t \sim T(\nu, \tilde{\mu}_{t+1}, \tilde{\delta})$$

$$\tilde{\mu}_{t+1} = \mu_{t+1} + (\lambda_t - \mu_t) e^{-\phi}$$

$$\tilde{\delta}^2 = \frac{\sigma^2}{2\phi} (1 - e^{-2\phi})$$

**Cox-Ingersoll-Ross model**

$$\lambda_i = \int_{t_i-1}^{t_i} \lambda_s ds, \quad i = 1, \dots, T$$

$$d\lambda_t = \alpha(\tilde{\mu}_t - \lambda_t) + \sigma \sqrt{\lambda_t} dW_t$$

$$\lambda_t \gamma | \lambda_t = \frac{Y_t}{2c}$$

where  $Y_t$  follows a non-central  $\chi^2$  with  $\frac{2c\mu_t}{\sigma^2}$  d.f. and non-centrality parameter  $2c\lambda_t e^{-\alpha T}$ .

**Application**

We apply our models in a sheeppox epidemic that happened in N. Evros Prefecture between December 1994 and December 1998, infecting 249 farms.

**Results**

We select the best model based on WAIC.

Model	WAIC
Gaussian OU	339
Student-t OU	346
CIR	377

The most important covariates using Gibbs variable selection: the number of farms infected the previous week, temperature and humidity.

**Future work**

- Perform Perquential Analysis and select the best-performing model based on scoring rules.
- Use Sequential Monte Carlo algorithms for on-line learning.
- Model the data using Random graph continuous time models.

## 8. Stochastic Epidemic Modeling of COVID-19 (Anastasios Apsemidis, AUEB):

STOCHASTIC EPIDEMIC MODELLING OF COVID-19  
Apsemidis and Demiris  
Athens University of Economics and Business

**The grand finale**

Day No.1197. The virus seems to have been settled.

- New wrong model
- Dual account for endemicity
- New information is added
- Vector field results
- Save the world (Season 3)

**The basic model**

$$dI_t \sim NB(\theta_t, \psi)$$

$$\theta_t = p_t \cdot \sum_{k=1}^{t-1} \pi_{t-k} C_k$$

$$C_t = \lambda_{t-h-1} S_{t-h-1} I_{t-h-1} / N$$

$$S_t = S_{t-1} - C_t - V_t + A - A \cdot S_{t-1} / N$$

$$I_t = \sum_{k=0}^{t-1} C_{t-k} - A \cdot I_{t-1} / N$$

$$\pi_s = \int_{s-0.5}^{s+0.5} \pi(t) dt$$

$$p_t = p_{(j)} \cdot I(t \in [t_j, t_{j+1} - 1])$$

$$\lambda_t = \lambda_{(j)} \cdot I(t \in [t_j, t_{j+1} - 1])$$

$p_{(j)} \sim N(\mu, 10^{-8})$

$$\mu = \frac{1}{b_{t+1} - I_t} \sum_{k=0}^{t-1} \sum_{l=1}^k p_{(l)} \frac{c_{t,k}}{\sum_{l=1}^k c_{t,l}}$$

$\lambda_{(j)} \sim \text{LogNormal}(0, 1)$

**Loss of immunity**

After  $t^*$  days,

$$r_t = (1 - p_t) \cdot \sum_{k=1}^{t-1} \pi_{t-k}^* \cdot C_k$$

recovered individual's return to susceptibility.

Thus, the  $S$ -state is updated via

$$S_t = S_{t-1} - C_t - V_t + A \cdot (1 - S_{t-1} / N) + r_{t-t^*}$$

The complete model accounts for endemicity through

- demography
- return to susceptibility

**Results**

**Infection rate predictions**

Use of 1 or 2 PC's of daily mobility  $m_t$ , either smoothed by the serial interval or not.

Post-processing of  $\lambda_t$ :

$$\mathbb{E}[\log(\lambda_t) | m_t] = g(m_t)$$

**Forms of  $g(\cdot)$**

- Linear regression
- Thin-plate smoothing
- Extreme-gradient boosted trees

**Final Remarks**

- Data from Greece, UK and USA
- Only publicly available data
- Computationally intensive training
- Different behaviour after 2021
- The end (?)

## 9. Model-based Indicators for in-play Basketball Data (Argyro Damoulaki, AUEB):

### Model-based Indicators for in-play Basketball Data

Argyro Damoulaki<sup>1</sup>, Ioannis Ntzoufras<sup>1</sup>, Konstantinos Pelechinis<sup>2</sup>  
<sup>1</sup>Athens University Economics Business, <sup>2</sup>University of Pittsburg

**Objectives**

- Objectives for today:
  - Aim of the work
  - Explanatory analysis
  - Metrics with Regression
  - Multinomial-based Metric
- Aim of the work**
- Measure the contribution of basketball players:
  - Use plus/minus performance indicators based on models.
  - Investigate and compare different evaluation models.
- Explanatory analysis**
- Final data:** sparse matrix 322.850 × 1437
  - 717 offenders
  - 717 defenders
  - points (response)
  - home\_off
  - season\_type
- Team Possessions Ratio (TPR)**
- TPR > 1: 93 Offenders + 86 Defenders
- (O-TPR): # team scoring possessions with player on-court vs. off-court
- (D-TPR): # zero-score possessions for opponent team with player on-court vs. off-court

**Metrics with Regression**

**Literature Proposal:**  
 Regularized Adjusted Plus/Minus (RAPM), Ridge with  $\lambda=2000$

**Ridge (with  $\lambda=\lambda_{min}$ ) => Problem:**  
 Results are influenced by LTPs since they appear high rated

**Lasso Regression:**

- Better automatic cutoff players
- Players kept: 184 offenders & 143 defenders
- Considerably better than ridge

**Low-time players (LTP):** < 200 minutes played

- Rosenbaum (2004): LTP threshold 250 minutes for 2 seasons
- Iardi and Barzilai (2008): LTP threshold 300 minutes for 1 season

**High time players: >175 MP**

**RAPM: No good fit**

**Ridge Binomial vs. Normal**

$Binomial_i = 0.02 \times Normal_i + \epsilon, \epsilon \sim N(0, 0.008^2)$  for j player

**Multinomial-based Metrics**

- Indirect implementation via 3 binomial models
- Screening Lasso: remain 362 offenders and 316 defenders
- 1 index: expected points in i possession

$Ept_i = P_1 + 2 \times P_2 + 3.01 \times P_3$

**External Validation Criteria**

External Validation Criteria	Normal	Multinomial
Top-3 NBA lineups	40%	53%
Positive TPR	32%	62%
Most-played lineups per team	47%	67%

**Offenders top 10 Normal vs. Multinomial**

**References**

[1] Rosenbaum, D.T., (2004) "Measuring How NBA Players Help Their Teams Win", 82games, <http://www.82games.com/comm30.htm>.  
 [2] Iardi, S. and Barzilai, A., (2007), "Adjusted Plus-Minus: An Idea Whose Time Has Come", 82games, <http://www.82games.com/iardi07.htm>.

## 10. The Macro Determinants of Global Tax Avoidance: a Systematic literature review (Athanasios Vasilakis, International Hellenic University):

### The Macro Determinants of Global Tax Avoidance: a Systematic literature review

Athanasios VASILAKIS<sup>†</sup>  
<sup>†</sup>Department of Economic Sciences, International Hellenic University

**Objectives**

- > To find the determinants of Base Erosion and Profit Shifting (BEPS) that has affected all countries by a number of macroeconomic parameters.
- > RQ1: To identify behaviors of macroeconomic dependent and independent determinants on the various economies.
- > RQ2: To provide a review of the macroeconomics of BEPS using the high-quality literature published between 2003 till 2022.

**The Macroeconomics of BEPS PRISMA (2020) flow diagram**

Independent and Dependent Variables	n	% of sources	Sources
1. Tax Rate	20	26%	Endler et al., 2022; Viscusi, 2020; von Brans et al., 2022; Alvarez-Martinez et al., 2022; Duffield et al., 2021; Bekker et al., 2020; Jansky & Palanik, 2019; Alexandrou et al., 2019; Ching, 2019; Coburn & Jansky, 2018; Dierker et al., 2017; Ching, 2016; Beer & Loepck, 2014; Dischinger & Riedel, 2011; De Mooij & Deruven, 2010; Betsendorfer et al., 2010; Wacziarg, 2009; Ching, 2009; Hines, 2009; Laatz & Mihov, 2007; Ching, 2003
2. GDP	18	24%	von Brans et al., 2022; Duffield et al., 2021; Wacziarg, 2020; Jansky & Palanik, 2019; Alexandrou et al., 2019; Ching, 2019; Coburn & Jansky, 2018; Dierker et al., 2017; Ching, 2016; Beer & Loepck, 2014; Dischinger & Riedel, 2011; De Mooij & Deruven, 2010; Betsendorfer et al., 2010; Wacziarg, 2009; Ching, 2009; Hines, 2009; Laatz & Mihov, 2007; Ching, 2003
3. Population	14	18%	von Brans et al., 2022; Duffield et al., 2021; Bekker et al., 2020; Jansky & Palanik, 2019; Alexandrou et al., 2019; Ching, 2019; Coburn & Jansky, 2018; Dierker et al., 2017; Ching, 2016; Beer & Loepck, 2014; Dischinger & Riedel, 2011; De Mooij & Deruven, 2010; Betsendorfer et al., 2010; Wacziarg, 2009; Ching, 2009; Hines, 2009; Laatz & Mihov, 2007; Ching, 2003
4. GDP Growth	13	17%	Alexandrou et al., 2020; De Mooij & Lin, 2020; Coburn & Jansky, 2019; Alexandrou et al., 2019; Ching, 2019; Coburn & Jansky, 2018; Alexandrou et al., 2018; Ching, 2016; Beer & Loepck, 2014; Dischinger & Riedel, 2011; Betsendorfer et al., 2010; Laatz & Mihov, 2007; Ching, 2003
5. GDP per capita	11	14%	Bekker et al., 2020; Alexandrou et al., 2020; Jansky & Palanik, 2019; Coburn & Jansky, 2018; Dierker et al., 2017; Ching, 2016; Beer & Loepck, 2014; Dischinger et al., 2011; De Mooij & Deruven, 2010; Dhammapala & Hines, 2009; Ching, 2003
6. Investments	10	13%	Viscusi, 2020; Alvarez-Martinez et al., 2022; von Brans et al., 2022; Duffield et al., 2021; Bekker et al., 2020; Jansky & Palanik, 2019; Betsendorfer et al., 2010; Dhammapala & Hines, 2009; Deruven et al., 2008; Laatz & Mihov, 2007
7. Corruption	6	8%	Duffield et al., 2021; Bekker et al., 2020; Alexandrou et al., 2020; Beer & Loepck, 2014; Dischinger & Riedel, 2011; Hainings & Lorenz, 2008
8. Population	5	7%	Wacziarg, 2020; De Mooij & Lin, 2020; David et al., 2017; Dischinger & Riedel, 2011; Dhammapala & Hines, 2009; Deruven et al., 2008
9. Real Exchange Percentage	5	7%	Duffield et al., 2021; Wacziarg, 2020; De Mooij & Lin, 2020; Ghisano et al., 2009; Ching, 2003
10. Openness	4	5%	Duffield et al., 2021; Coburn & Jansky, 2018; Ghisano et al., 2009; Deruven et al., 2008
11. Employment	4	5%	Bekker et al., 2020; De Mooij & Lin, 2020; Dischinger & Riedel, 2011; De Mooij & Deruven, 2010
12. Inflation	4	5%	Duffield et al., 2021; Alexander et al., 2020; Coburn & Jansky, 2018; Dierker, 2010
13. Tax Burden	4	5%	Viscusi, 2022; Alvarez-Martinez, 2022; Jansky & Palanik, 2019; Cerrilli et al., 2016

**Dependent Determinant**

Dependent Determinant	No. of Sources	% of Sources
1. Tax	4	13%
2. Tax	4	13%
3. Tax Base	3	10%
4. Transfer	3	10%
5. Pricing	3	10%
Study country	20	67%
multi country	30	

**DISCUSSION**

**Answering RQ 1**, in this study it is more than clear that BEPS has significantly affected all countries around the world. There are several determinants that influence BEPS but the dependent variables that reflect BEPS also affect the overall economic performance of the states.

**Answering RQ 2**, this paper provides a review of the macroeconomics of BEPS using the high-quality literature published between 2003 till 2022. It takes into account mainly macroeconomic Tax impacts on BEPS. The changes in macroeconomic dependent and independent determinants and containment tax policies have been considerations in the decision-making process of states.

**CONTRIBUTION**

- The BEPS literature relating to macroeconomics have been placed in a new clear and understandable context.
- This study has identified various determinants of BEPS by tabulating the relationships, demonstrating their respective influences.
- Explore further the relationships to the extent countries of Origin or Host countries can recover their macroeconomic loss.

**Acknowledgements**

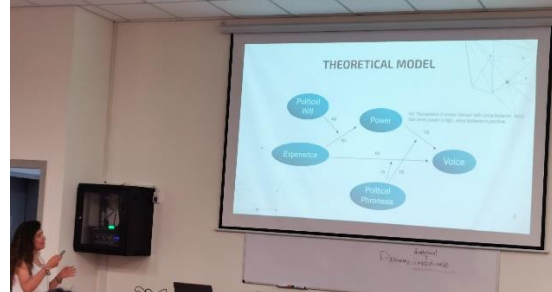
I would like to thank libraries and librarians of International Hellenic University (Serres), Bank of Greece and Economic University of Athens for their support. Moreover, I would like to thank Assistant Professor Vassilios VLACHOS and the anonymous reviewers of 24rd Annual Conference on Finance and Accounting Prague 1-2 June 2023 for their fruitful comments.

Η πρώτη ημέρα του Workshop ολοκληρώθηκε στο Αμφιθέατρο (Όροφος -1), του Νέου Κτιρίου του ΟΠΑ (*Troias 2 and Spetson*), με τις κάτωθι 3 Προφορικές Παρουσιάσεις, με προεδρεύουσα την Chiara Saccon (*Universita Ca Foscari, Venezia*):



1. Irina Rodriguez De La Flor Demarcos (*University Of Alcalá*): [Inner Knowledge, A New Tool For Businesses](#)

2. Stoumpou Dimitra (*AUEB*) : [Exploring the Interplay of Individual Experience, Perceived Power, and Political Will: Insights into Voice Behavior among Managers and Leaders:](#)



3. Rizwan Ahmad (*Universita Ca Foscari, Venezia*): [Big Data and Hospital Sustainable Performance: Unpacking the Role of Green Suppliers Collaboration, Resilience and Commitment:](#)



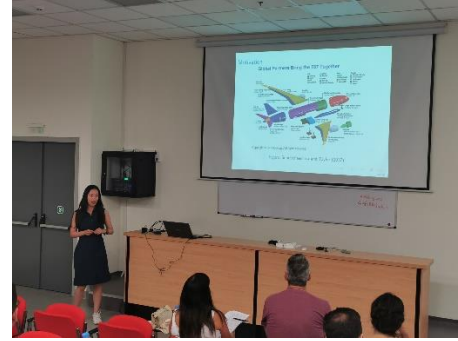
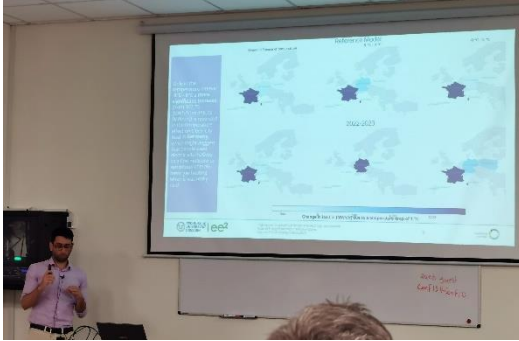
Μετά το πέρας των Παρουσιάσεων, οι συμμετέχοντες αποχαιρετίστηκαν και ανανέωσαν το ραντεβού τους, για την επόμενη ημέρα:



Η δεύτερη ημέρα (8 Ιουνίου 2023) ξεκίνησε δυναμικά, με προεδρεύων των τριών πρώτων Προφορικών Παρουσιάσεων, τον Ιωάννη Ντζούφρα (*Athens University of Economics and Business*):

1. Glynos Dimitrios (Technische Universität of Dresden) : [Effects of temperature on gas and electricity consumption in European countries: A high-resolution data analysis](#)

2. Yao Zhixiao (Technische Universität of Dresden): [Speaking a Common Technical Language: ISO Membership and Non-tariff Trade Barriers](#)



3. Anna Nalpantidi (AUUEB): [Model Based Clustering for Spatial Data:](#)



Ακολούθησε ένα σύντομο διάλειμμα, κατά τη διάρκεια του οποίου, έγινε και η λήψη των αναμνηστικών φωτογραφιών των συμμετεχόντων του Workshop:

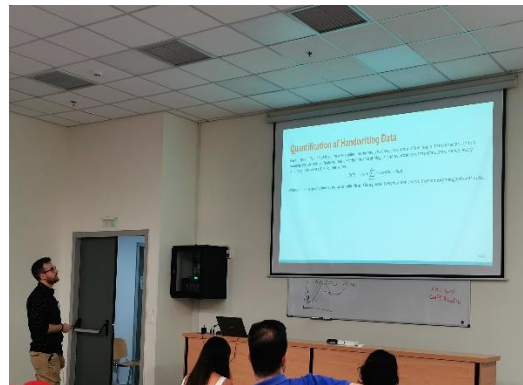
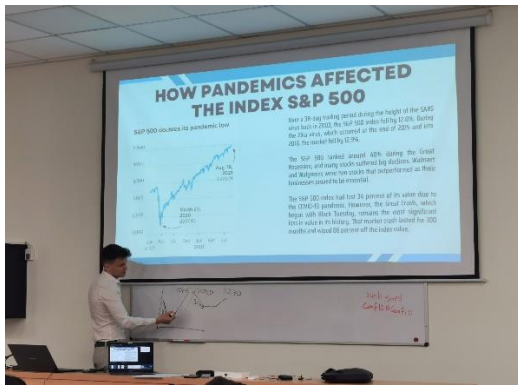




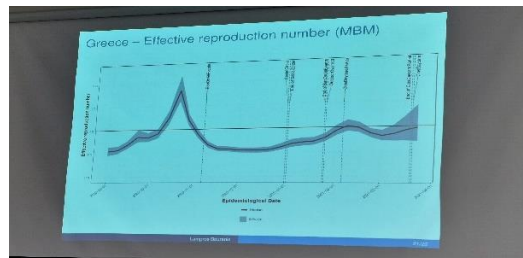
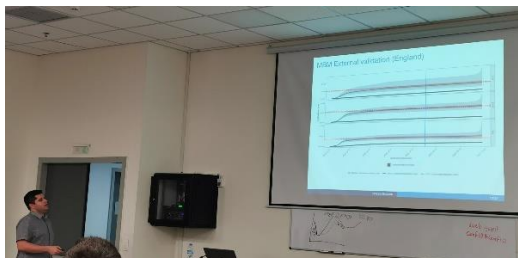


Η τελευταία Συνεδρία του Workshop, με προεδρεύουσα την Claudia Tarantola (*Università di Pavia*), περιλάμβανε τις εξής Προφορικές Παρουσιάσεις:

1. Kalamen Kristian (*University of Bratislava*) : [Pandemic economic crises](#)
2. Lampis Tzai (*AUEB and UNIL*): [Statistical Examination Handwriting Evidence in Forensic Science](#):



3. Lambros Bouranis (*AUEB*) [Bayesian analysis of diffusion-driven multi-type epidemic models with application to COVID-19](#):



Τη λήξη του Workshop σηματοδότησε το μεσημεριανό γεύμα, στο Εστιατόριο του Κεντρικού Κτιρίου του ΟΠΑ, το οποίο και χαρακτηρίζονταν από πολλά χαμόγελα και υγιεινή διατροφή:



Το ραντεβού ανανεώθηκε για το βράδυ της ίδιας ημέρας, στο δείπνο που δόθηκε στους συμμετέχοντες που έκαναν Προφορική Παρουσίαση. Απολαύσαμε το ηλιοβασίλεμα, βγάλαμε πολλές φωτογραφίες και δημιουργήσαμε αναμνήσεις, που θα μας συνοδεύουν για μία ζωή:













Καθώς το Workshop έφτασε στο τέλος του, συναισθήματα χαράς μας κατέκλυσαν. Συναισθήματα που συνέκλιναν στον αμοιβαίο σεβασμό προς την ερευνητική δουλειά των συναδέλφων μας, την εμπαθή ανατροφοδότηση, τις εποικοδομητικές συνομιλίες, την εκτίμηση προς τη δικτύωση, αλλά και την απόλυτη ικανοποίηση!

Ευχαριστούμε όλους όσους συμμετείχαν στο Workshop, γι' αυτή την αξέχαστη εμπειρία, καθώς και το [Higher Education and Research in Management of European Universities \(HERMES\)](#), που έκανε πραγματικότητα τη συγκεκριμένη διοργάνωση! Ευχόμαστε, από καρδιάς, αυτή να είναι μόνο η απαρχή παρόμοιων πολυπολιτισμικών δρώμενων, στο Πανεπιστήμιό μας και το Τμήμα Στατιστικής!

Για περισσότερες πληροφορίες, για το International HERMES Ph.D. Workshop 2023: "Data Science in Business", παρακαλούμε όπως ανατρέξετε στην επίσημη Ιστοσελίδα του Workshop:

<https://sites.google.com/view/aueb-phd-hermes-workshop-2023/home>