Estatística, Econometria, Ciência de Dados no INSPER

Hedibert F. Lopes

Professor titular de estatística e econometria

Coordenador do Núcleo de Ciências de Dados e Decisão

www.hedibert.org

30 setembro de 2024

80+ artigos científicos publicados em revistas internacionais de alto impacto:

- Journal of the American Statistical Association
- Journal of Computational and Graphical Statistics
- Journal of Business and Economic Statistics
- American Economic Review

Bolsista de Produtividade em Pesquisa do CNPq nível 1C

10 mil citações no Google Scholar

***MCMC: Stochastic Simulation for Bayesian Inference* -** 4575 citações

Primeiro brasileiro eleito Fellow da International Society for Bayesian Analysis (2022)

Doutores em Estatística e Economia que orientei entre 2017 e 2024

INSPER

Rafael Campello de Alcantara
Post-doc at UT Austin
D.Sc. in Business Economics (Fev-2024)
Cutoff-aware BART for estimating Heterogeneous Treatment Effects in Regression Discontinuity Designs

Igor Ferreira Batista Martins
Post-doc in Econometrics at Örebro, Sweden
D.Sc. in Business Economics (Feb-2024)
Essays in Bayesian Financial Econometrics

Bruno do Prado Costa Levy
Itaú Asset Management
D.Sc. in Business Economics (2021)
Dynamic ordering learning in multivariate forecasting

USP

Renata Tavanielli
D.Sc. in Statistics (2024)
Cholesky-based dynamic copula modeling

Henrique Bolfarine
Professor UT Austin
D.Sc. in Statistics (2021)
DSS in Gaussian Linear Factor Analysis

Helton Graziadei
Professor at UFPR
D.Sc. in Statistics (2020)
Some Bayesian generalizations of the INAR model

Paloma Waisman Uribe
Data Science Manager at Nubank
D.Sc. in Statistics (2017)
Dynamic sparsity on time-varying Cholesky-based covariance matrices

Doutores em Estatística, Economia e Engenharia Elétrica que orientei entre 2003 e 2014

Chicago Booth

Samir Warty
Senior Data Scientist, Analysis Group

Maria Paula Rios
VP of Inovation and Digital Transformation, Alianza Team

Paolo Bonomolo
Senior economist, Research Dept, The Nederlansche Bank
Universita degli studi di Pavia

Bruno Lund
Chief Investiment Officer, Ecoagro
EPGE-FGV

Edison Tito
Assistante Professor, IME-UERJ
DEE-PUC/RJ

IM-UFRJ

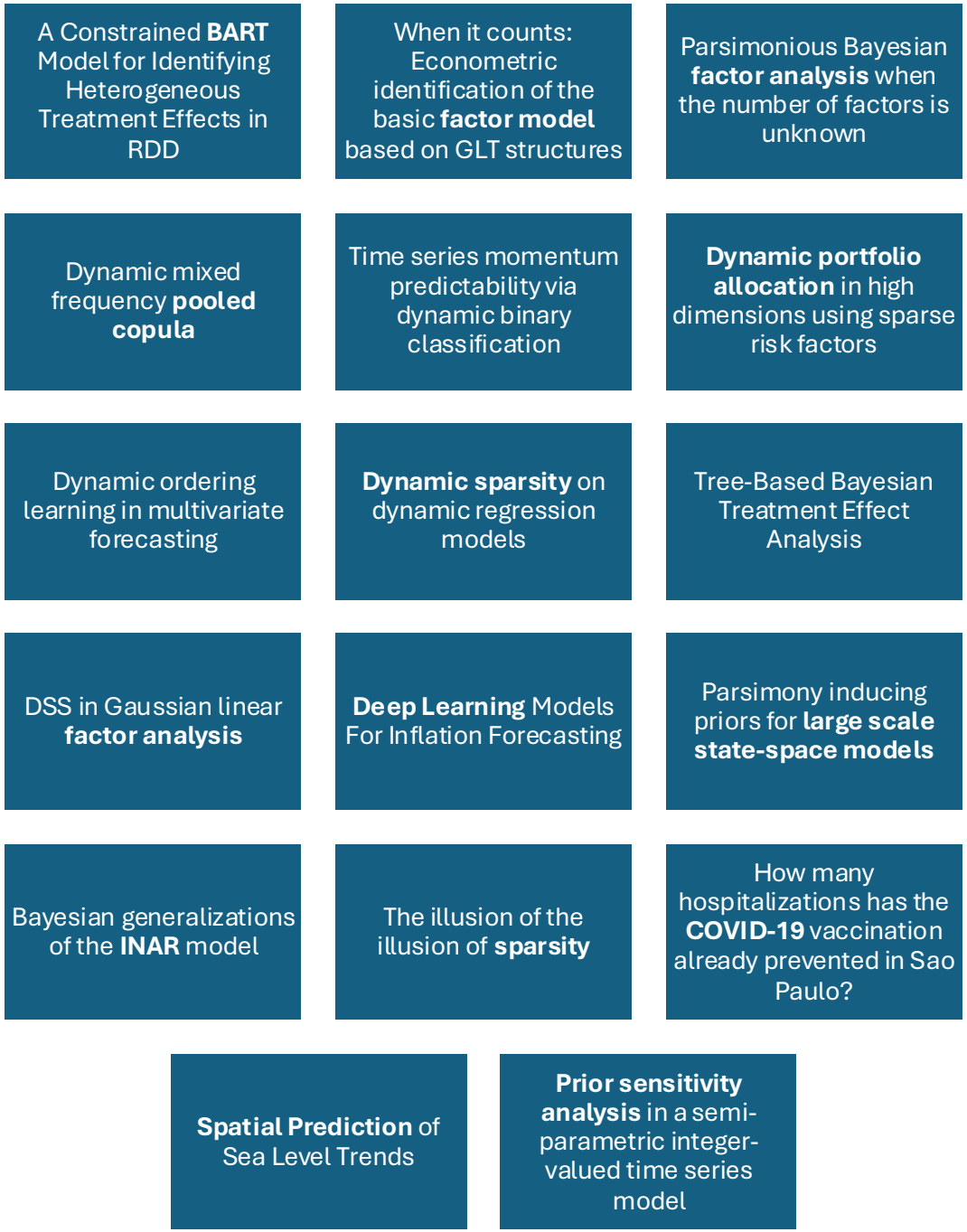
Esther Salazar
Senior Mathematical Statistician, FDA

Fernando F. Nascimento
Assistant Professor, UFPI

Ralph S. Silva
Associate Professor, IM-UFRJ

Carlos Abanto-Valle
Associate Professor, IM-UFRJ

Cibele Behrens
Quatro Consultoria Economica Ltda



Scalable semi-parametric inference for the means of **heavy-tailed** distributions

Bayesian semi-parametric **MSSV** models

Walk on the wild side: **Multiplicative sunspots** and temporarily unstable path

Efficient sampling for Gaussian linear regression with arbitrary priors

Particle learning for Bayesian **semi-parametric SV** model

Dynamic models

Bayesian hypothesis testing: Redux

On the long run volatility of stocks: time-varying predictive systems

Bayesian factor model **shrinkage** for linear IV regression with many instruments

Sequential Bayesian learning for SV with **variance-gamma jumps** in returns

Efficient Bayesian inference for **multivariate factor SV models**

Cholesky **realized stochastic volatility** model

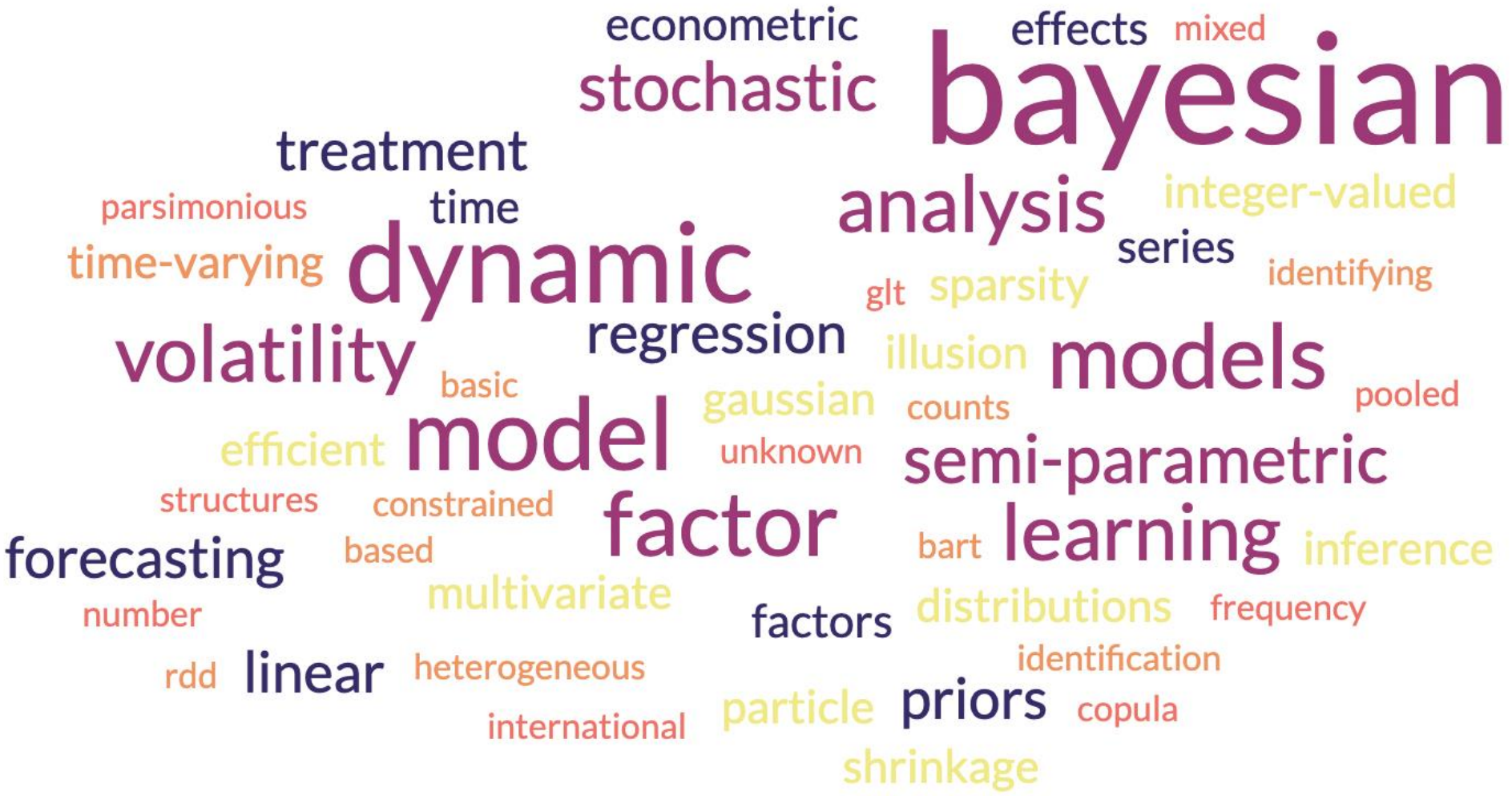
Particle learning for **fat-tailed** distributions

Time-varying extreme pattern with dynamic models

Bayesian instrumental variables: likelihoods and priors

Treatment effects: a Bayesian perspective, Econometric Reviews

Modern Bayesian Factor Analysis


econometric
effects
mixed
stochastic
bayesian
treatment
analysis
integer-valued
parsimonious
time
time-varying
dynamic
series
identifying
glt
sparsity
volatility
regression
illusion
models
basic
gaussian
pooled
efficient
model
counts
unknown
semi-parametric
structures
constrained
factor
learning
inference
forecasting
based
bart
multivariate
number
factors
distributions
frequency
identification
rdd
linear
heterogeneous
particle
priors
copula
international
shrinkage

Bayesianity x frequentism

Overfitting

Prior predictive: “Bayesian integration over parameters automatically and elegantly penalizes complexity.” (Bishop & Bishop, 2024, page 57).

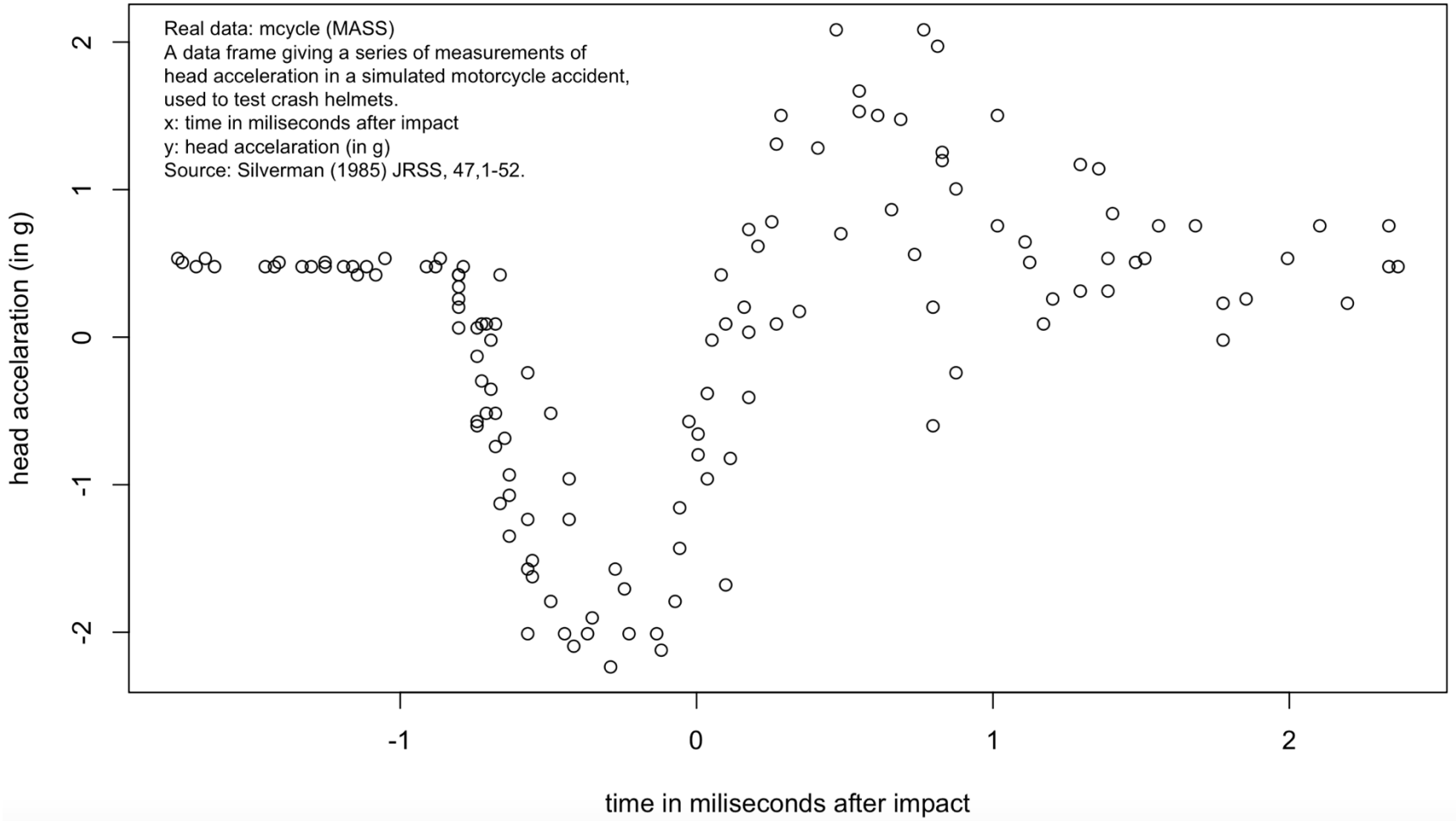
Estimation vs point-null hypothesis testing

Unit root: Straightforward posterior inference vs unnecessary Brownian motion for point-null hypothesis testing. Hundreds and hundreds of pointless scientific papers.

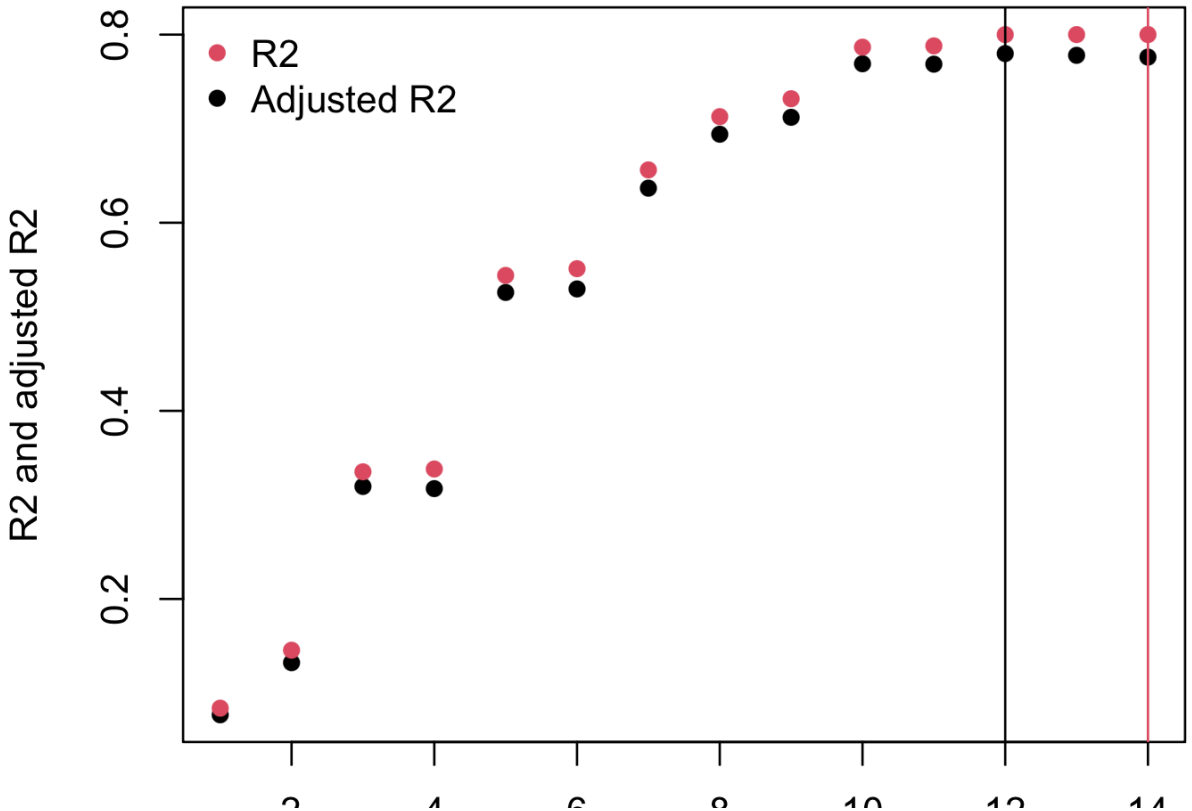
Normal dynamic linear modeling

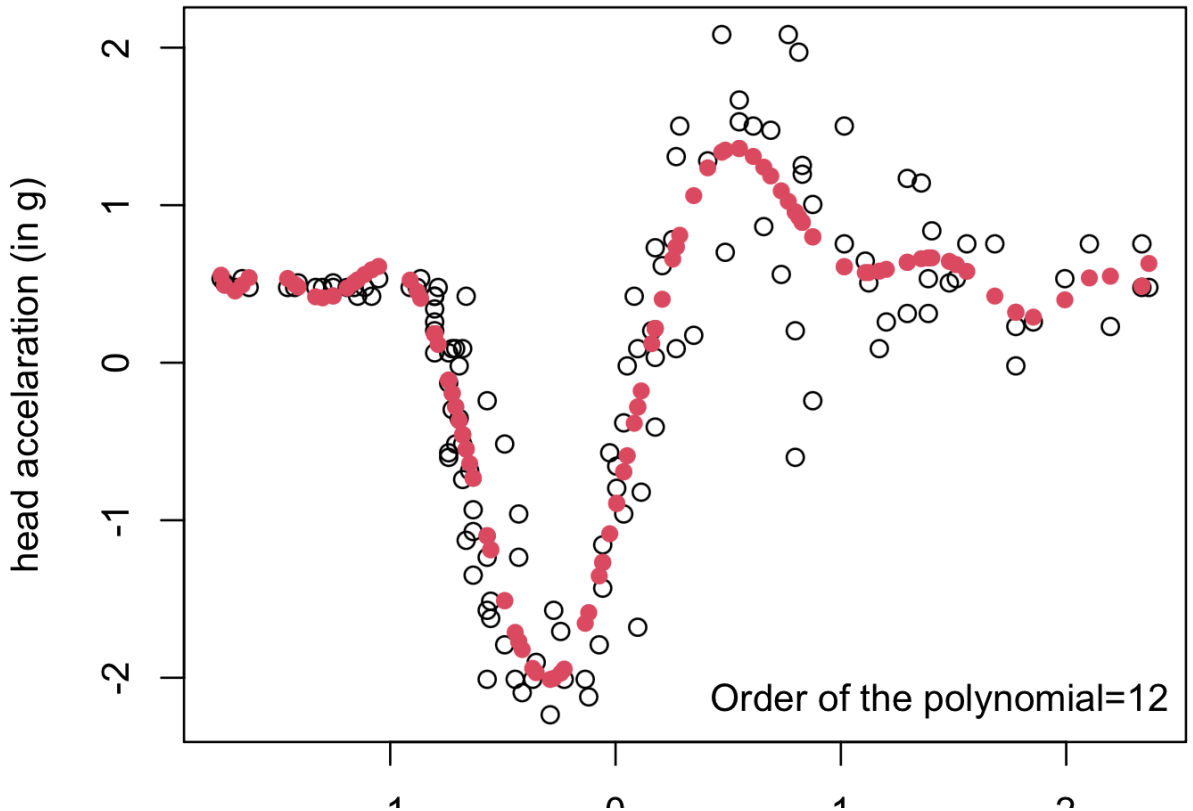
Uncertainty quantification (UQ): Two-step kalman-filter & parameter estimation vs fully Bayesian uncertainty account

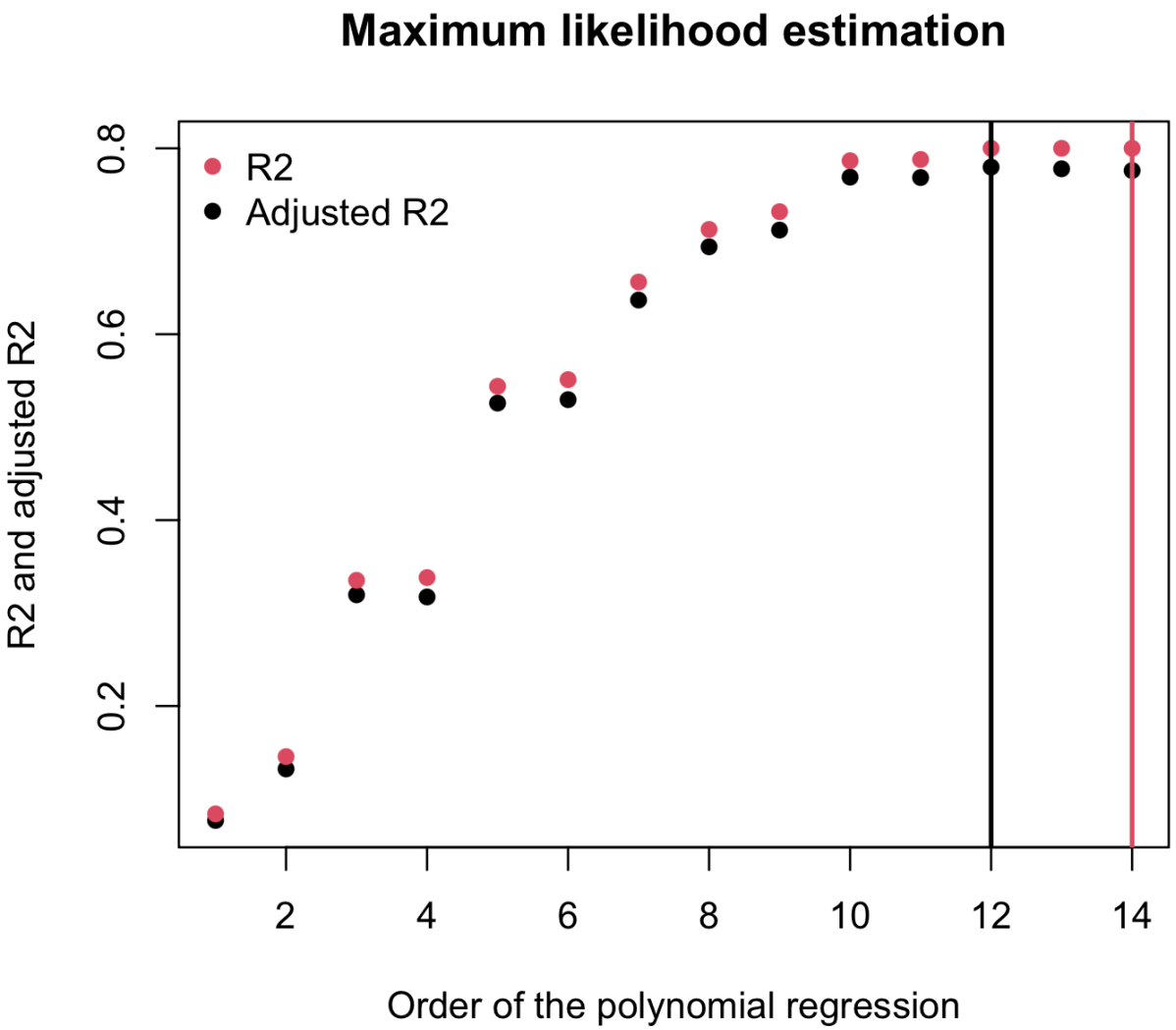
Motorcycle helmets

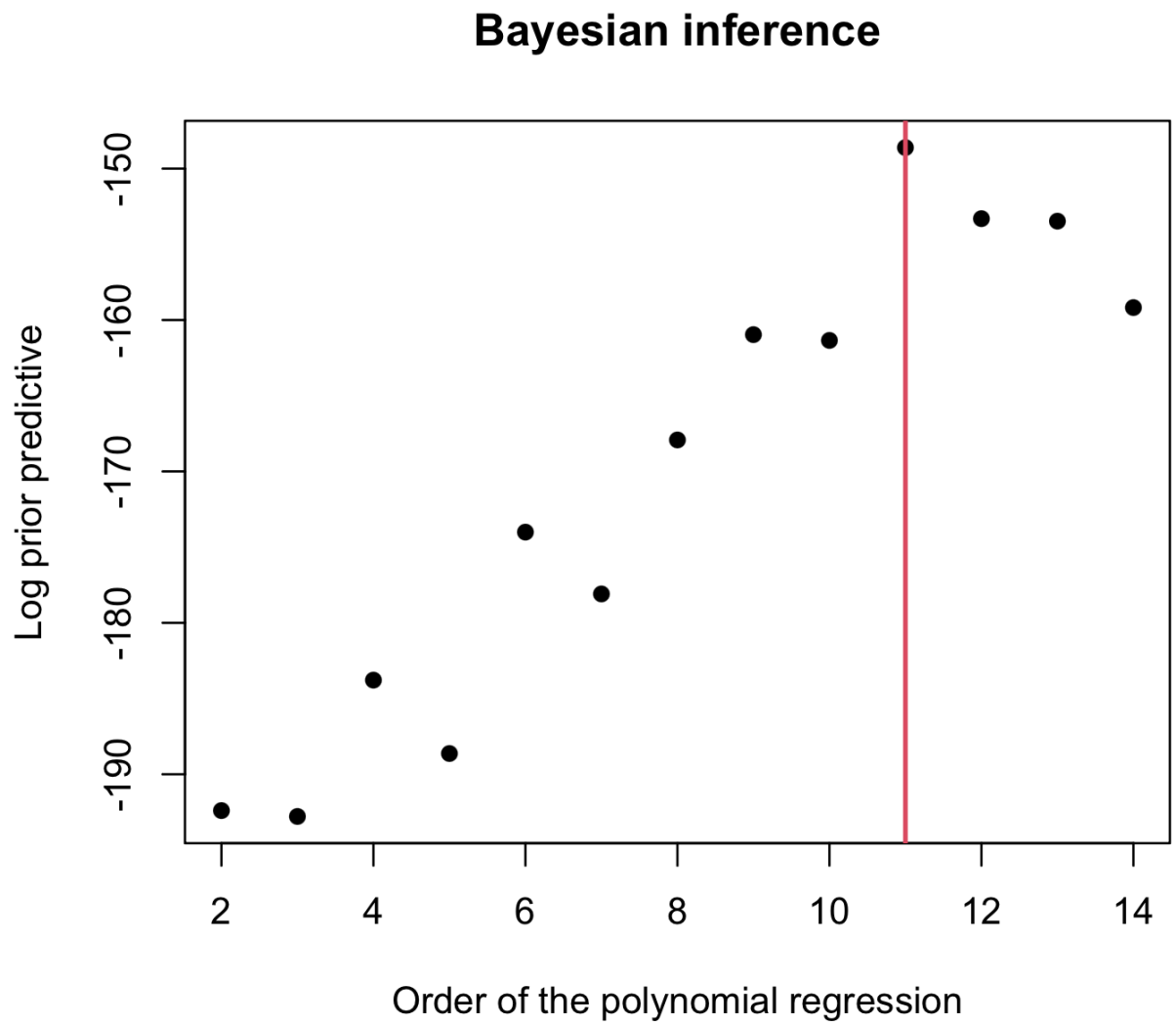


Polynomial regression

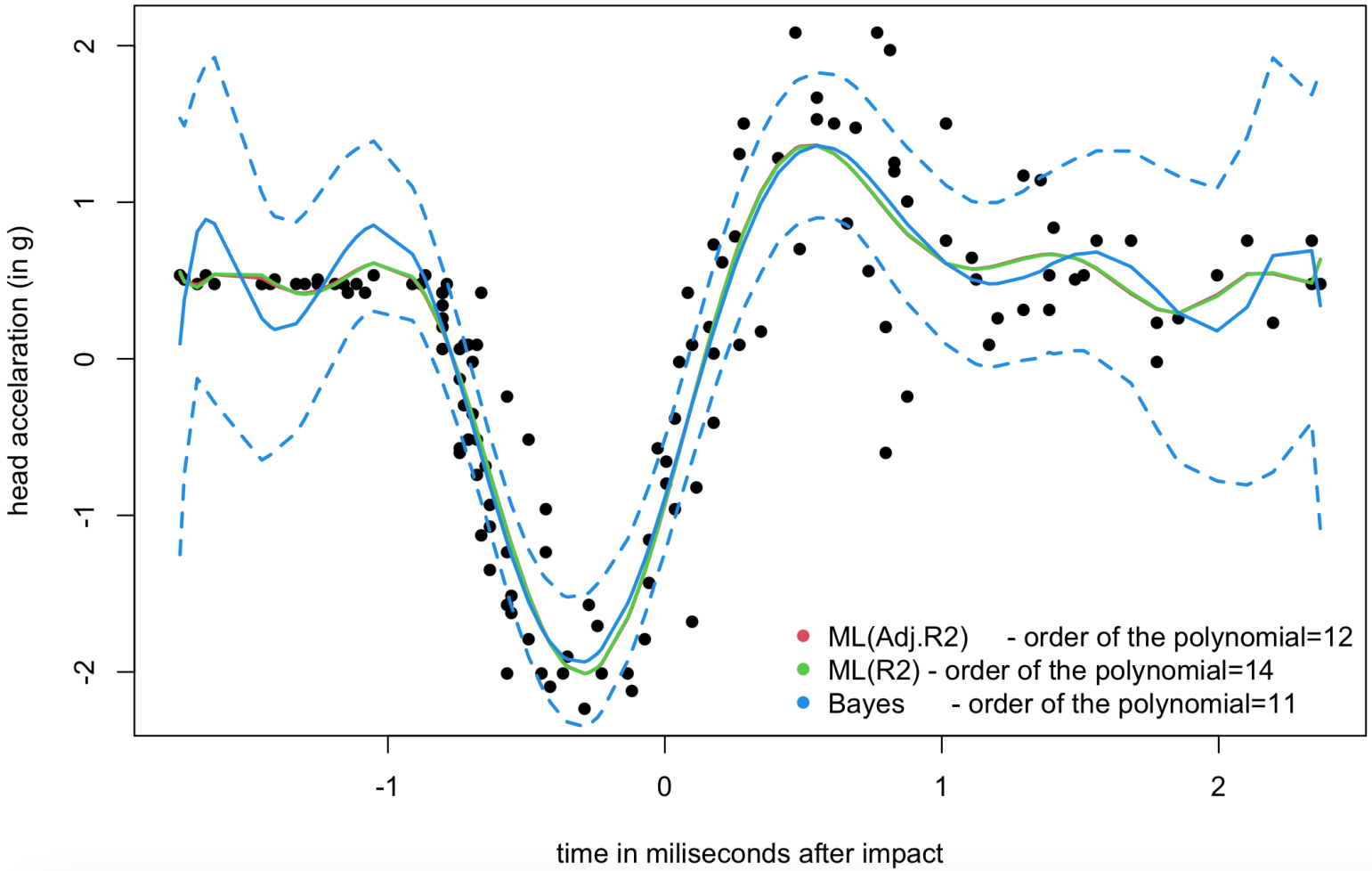





Maximum likelihood estimation
R2
Adjusted R2
R2 and adjusted R2
0.2
0.4
0.6
0.8
2
4
6
8
10
12
14
Order of the polynomial regression


Bayesian inference
Log prior predictive
-150
-160
-170
-180
-190
2
4
6
8
10
12
14
Order of the polynomial regression

Posterior model probability for order 11 = 98.3%



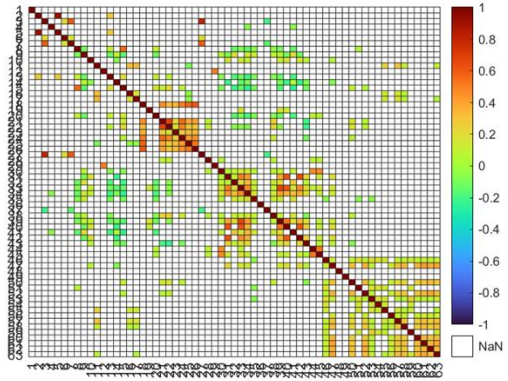
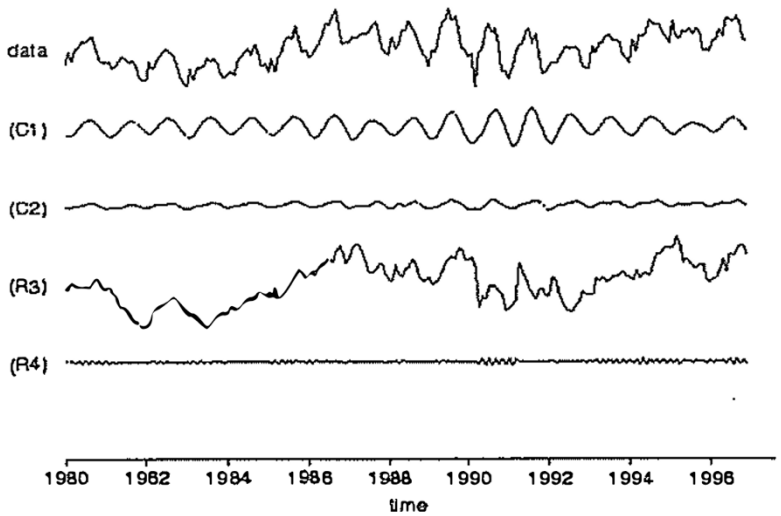
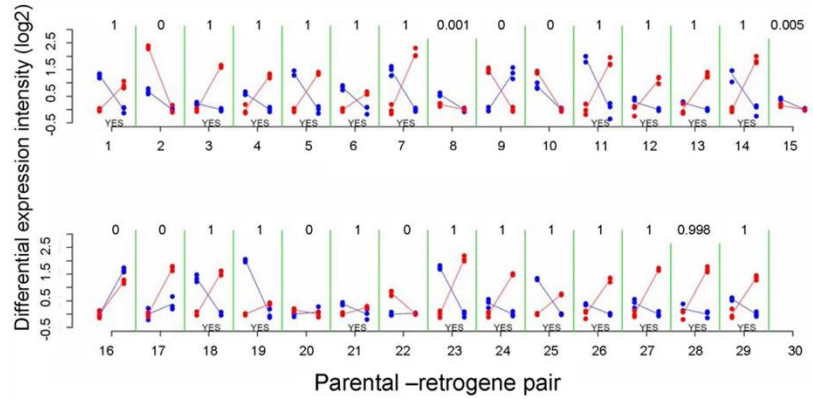
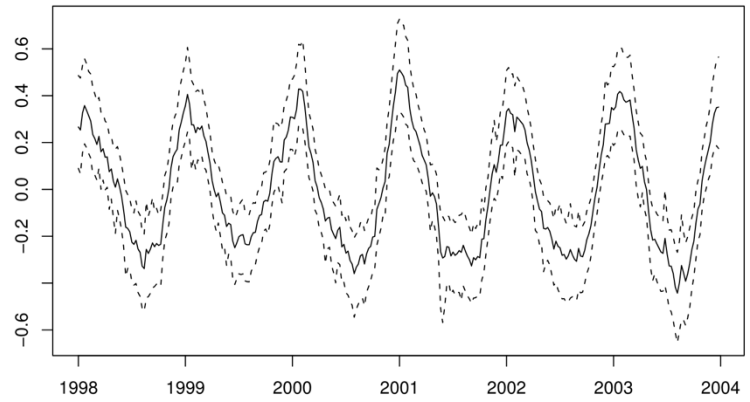


Figure 7: NYSE data; estimated marginal correlation matrix $\mathrm{E}(\mathbf{\Omega}^{\star}|\mathbf{y})$, where $\Omega^{\star}_{i\ell} = \mathrm{Corr}(\Lambda_{i1}f_{1t})(y_{\ell t} - \Lambda_{\ell 1}f_{1t}))$.







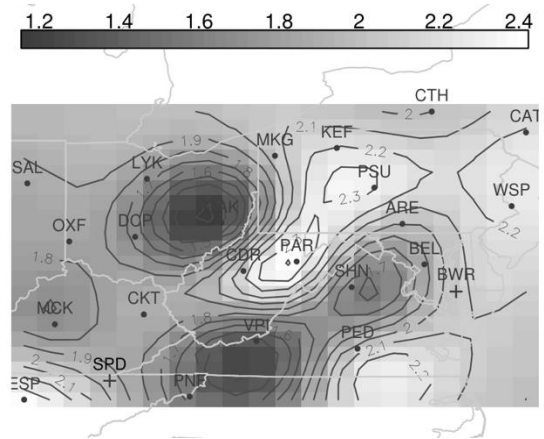
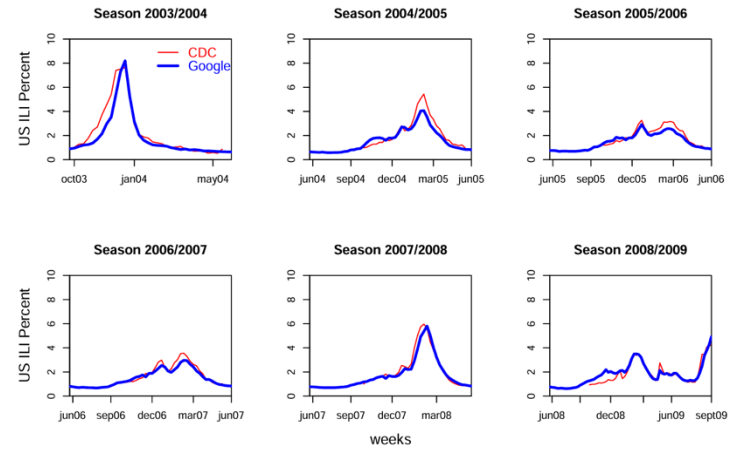


Figure 2: Google Flu Trends estimated ILI percentages (dashed line) and CDC ILI Surveillance percentages (solid line) for the United States, from June 2003 until September 2009. Separate plots correspond to separate influenza years, with each new influenza season starting in autumn, and ending in spring. Note that CDC did not used to produce ILI reports during summers before 2009, and thus no solid line appears during summer months prior to 2009.



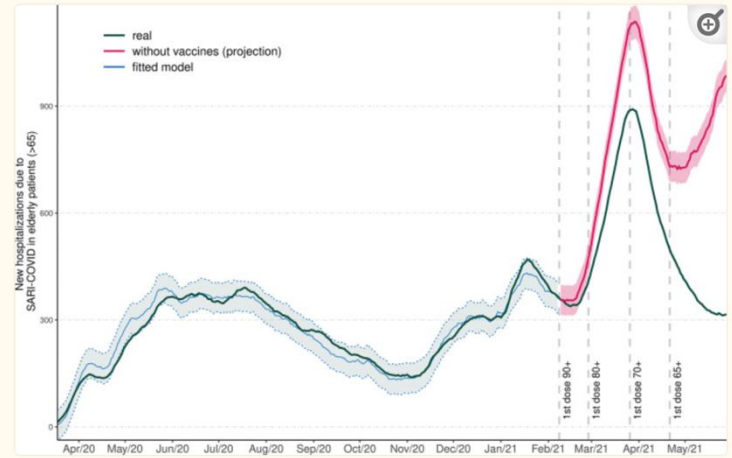


Figure 1

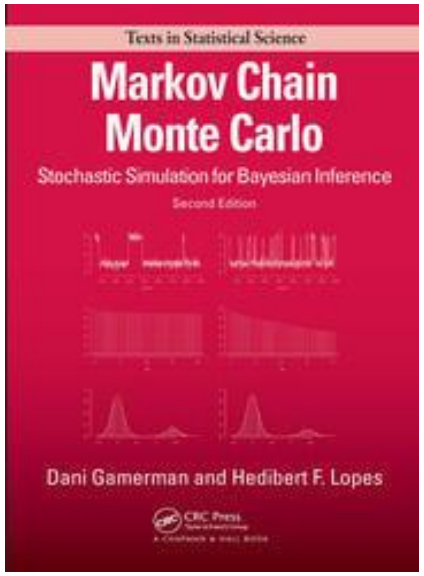
Number of hospitalizations due to SARI-COVID in patients aged >65 years (dark green), fitted pre-vaccination model (blue), and estimated counterfactual curve for the setting without vaccines (red).

Nosso time de estatísticos

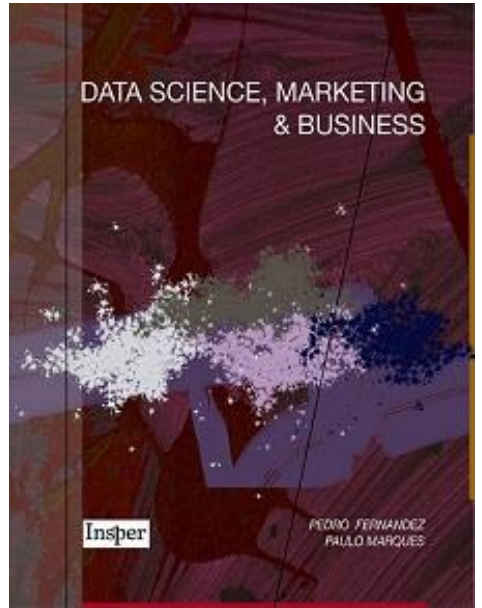
Rinaldo Artes
Professor Titular

Adriana Bortoluzzo
Professora Titular

Paulo Marques
Professor Assistente



Jovens talentos
Tiago Mendonça
Yasmin Cavalieri
Julio Trecenti
Magno Severino



Rinaldo Artes

Entrepreneurship in times of economic stress: unraveling the U-shaped relationship between the internality of causal attributions and growth. **Journal of Small Business and Enterprise Development**, 2024.

The impact of gestational weight gain on fetal and neonatal outcomes: the Araraquara Cohort Study. **BMC Pregnancy and Childbirth**, 2024.

The circular quantile residual. **Computational Statistics and Data Analysis**, 2023.

Estimating credit and profit scoring of a Brazilian credit union with logistic regression and machine-learning techniques. **RAUSP Management Journal**, 2019.

Adriana Bortoluzzo

Multichannel relational communication strategy: does one-sized strategy fit all customers? **European Journal of Marketing**, 2024.

Direção do mercado acionário impacta o alfa de fundos? **Revista Brasileira de Finanças**, 2023.

Forecasting accuracy of industrial sales with endogeneity in firm-level data. **International Journal of Business Marketing and Management**, 2021.

Disparity in the access to kidney transplantation for sensitized patients in the state of Sao Paulo-Brazil. **Transplant Immunology**, 2021.

Paulo C. Marques F.

Probabilistic Nearest Neighbors Classification. **Entropy**, 2024.

Confidence intervals for the random forest generalization error. **Pattern Recognition Letters**, 2022.

Prior Sensitivity Analysis in a Semi-Parametric Integer-Valued Time Series Model. **Entropy**, v. 22, p. 69, 2020. Citações:4|3

Bayesian generalizations of the integer-valued autoregressive model. **Journal of Applied Statistics** 2020.

Programas de Pós-Graduação do INSPER

Doutorado em Economia dos Negócios

Estratégia & Marketing

Microeconomia

Macroeconomia & Finanças

Doutorado Profisssional em Administração

Estratégias em organizações privadas

Estratégias em organizações públicas e do terceiro setor.

Mestrado Profissional em Economia

Mestrado Profissional em Administração

Mestrado Profissional em Políticas Públicas

Programa Avançado em Data Science e Decisão

Mestrado Profissional em Ciências de Dados e Decisão (MPCDD)

Em elaboração

Programa Avançado em Data Science e Decisão

1° trimestre

- Aprendizagem Estatística de Máquina I
- Computação para Ciência de Dados

2° trimestre

- Aprendizagem Estatística de Máquina II
- Big Data e Computação em Nuvem

3° trimestre

- Prática Avançada de Data Science e Visualization
- Métodos de Design para Ciência de Dados
- Data Science Deploy

4° trimestre

- Financial analytics
- Marketing analytics

5° trimestre

- Projeto de encerramento (Capstone) – Análises iniciais
- Clínicas

6° trimestre

- Projeto Capstone: conclusão do curso
- Clínicas (Design)

Mestrado Profissional em Ciências de Dados e Decisão (MPCDD) – Em elaboração

Disciplinas obrigatórias

1. Inferência Estatística
2. Probabilidade e Processos Estocásticos
3. Processamento de Dados em Escala
4. Sistemas de base de dados distribuídos
5. Introdução a Aprendizagem Estatística de Máquina

Concentração: Modelagem Estatística de Máquina

6. Estatística Bayesiana
7. Estatística Computational
8. Análise de Séries Temporais and/or Inferencia Causal

Concentração: Computação e Decision Analytics

6. Otimização
7. Aprendizagem Estatística de Máquina Avançada
8. Sistemas de Administração Base de Dados Avançado

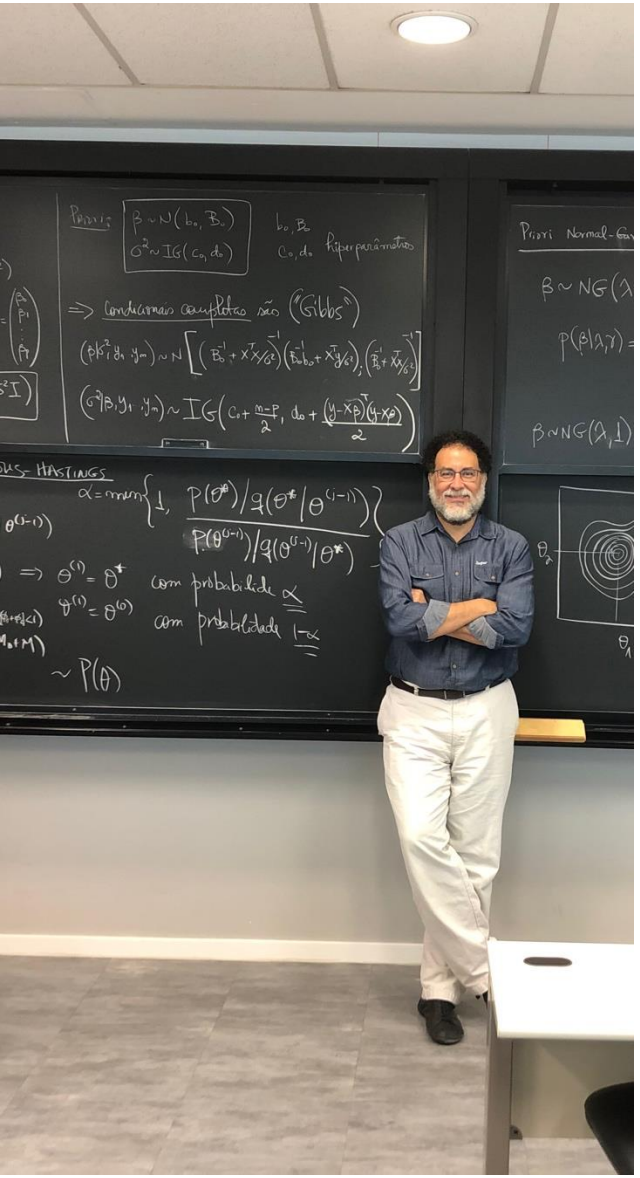
Inspiração: MS Data Science, Analytics and Engineering, Arizona State University

Doutorado Direto & Pós-doutorado no Insper

Pesquisador principal em 2 Projetos Temáticos FAPESP

Uma bolsa de doutorado direto (DD)

Duas bolsas de pós-doutorado (PD)


Priori:
Condicionais completas são (Gibbs)
com probabilidade α
com probabilidade 1−α